(19) World Intellectual Property Organization International Bureau



(43) International Publication Date 4 July 2002 (04.07.2002)

PCT

(10) International Publication Number WO 02/052015 A2

- (51) International Patent Classification⁷: C12N 15/51, 15/40, C12Q 1/68, 1/70, C12N 5/10, 7/04, 15/85
- (21) International Application Number: PCT/CA01/01843
- (22) International Filing Date:

20 December 2001 (20.12.2001)

(25) Filing Language:

English

(26) Publication Language:

English

- (30) Priority Data: 60/257,857
- 22 December 2000 (22.12.2000) US
- (71) Applicant (for all designated States except US): BOEHRINGER INGELHEIM (CANADA) LTD. [CA/CA]; 2100 Cunard Street, Laval, Québec H7S. 2G5 (CA).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): KUKOLJ, George [CA/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA). PAUSE, Arnim [DE/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).

- (74) Agent: BERNIER, Louise, G.; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

 without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: SELF-REPLICATING RNA MOLECULE FROM HEPATITIS C VIRUS

	CLONE APGK-12	AMINO ACID SUBSTITUTIONS								
		EMCV HCV NS2→5B					3.HCA			
	IRES	NeoR:	IRES	NS2	NS3	4A	NS4B	NS5A	NS5B	UTR
T7 ctalyg	G (nt1) SEQ ID NO 1									
	A (ntt) SEQ ID NO 24			-	•	•	-	•	-	
88 ch/µg	R3 rep A(nt1) SEQ ID NO 25				R(1135)K S(1560)G	K(1891)R	•	T(1993)A G(2042)C L(2155)P P(2166)L		
2000000ctu/µg	G(nt1) SEQ ID NO 7				R(1135)K S(1560)G	K(1891)R	-	T(1893)A G(2042)C L(2155)P P(2166)L		

(57) Abstract: A unique HCV RNA molecule is provided having an enhanced efficiency of establishing cell culture replication. Novel adaptive mutations have been identified within the HCV non-structural region that improves the efficiency of establishing persistently replicating HCV RNA in cell culture. This self-replicating polynucleotide molecule contains, contrary to all previous reports, a 5'-NTR that can be either an A as an alternative to the G already disclosed and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule. The G-->A mutation gives rise to HCV RNA molecules that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, possess greater efficiency of transduction and/or replication. These RNA molecules when transfected in a cell line are useful for evaluating potential inhibitors of HCV replication.



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SELF-REPLICATING RNA MOLECULE FROM HEPATITIS C VIRUS

FIELD OF THE INVENTION

The present invention relates generally to a HCV RNA molecule that self-replicates in appropriate cell lines, particularly to a self-replicating HCV RNA construct having an enhanced efficiency of establishing cell culture replication.

BACKGROUND OF THE INVENTION

disease have yet to be established.

Hepatitis C virus (HCV) is the major etiological agent of post-transfusion and
community-acquired non-A non-B hepatitis worldwide. It is estimated that over 200
million people worldwide are infected by the virus. A high percentage of carriers
become chronically infected and many progress to chronic liver disease, so called
chronic hepatitis C. This group is in turn at high risk for serious liver disease such as
liver cirrhosis, hepatocellular carcinoma and terminal liver disease leading to death.

The mechanism by which HCV establishes viral persistence and causes a high rate
of chronic liver disease has not been thoroughly elucidated. It is not known how
HCV interacts with and evades the host immune system. In addition, the roles of
cellular and humoral immune responses in protection against HCV infection and

20

25

Various clinical studies have been conducted with the goal of identifying pharmaceutical compounds capable of effectively treating HCV infection in patients afflicted with chronic hepatitis C. These studies have involved the use of interferonalpha, alone and in combination with other antiviral agents such as ribavirin. Such studies have shown that a substantial number of the participants do not respond to these therapies, and of those that do respond favorably, a large proportion were found to relapse after termination of treatment. To date there are no broadly effective antiviral compounds for treatment of HCV infection.

30 HCV is an enveloped positive strand RNA virus in the *Flaviviridae* family. The single strand HCV RNA genome is of positive polarity and comprises one open reading frame (ORF) of approximately 9600 nucleotides in length, which encodes a linear polyprotein of approx. 3010 amino acids. In infected cells, this polyprotein is cleaved at multiple sites by cellular and viral proteases to produce structural and non-structural (NS) proteins. The structural proteins (C, E1, E2 and E2-p7) comprise

5

10

15

20

25

30

35

2

polypeptides that constitute the virus particle (Hijikata et al., 1991; Grakoui et al., 1993(a)). The non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, NS5B) encode for enzymes or accessory factors that catalyze and regulate the replication of the HCV RNA genome. Processing of the structural proteins is catalyzed by host cell proteases (Hijikata et al., 1991). The generation of the mature non-structural proteins is catalyzed by two virally encoded proteases. The first is the NS2/3 zincdependent metalloprotease which auto-catalyses the release of the NS3 protein from the polyprotein. The released NS3 contains a N-terminal serine protease domain (Grakoui et al., 1993(b); Hijikata et al., 1993) and catalyzes the remaining cleavages from the polyprotein. The released NS4A protein has at least two roles. First, forming a stable complex with NS3 protein and assisting in the membrane localization of the NS3/NS4A complex (Kim et al., Arch Virol. 1999, 144: 329-343) and second, acting as a cofactor for NS3 protease activity. This membraneassociated complex, in turn catalyzes the cleavage of the remaining sites on the polyprotein, thus effecting the release of NS4B, NS5A and NS5B (Bartenschlager et al., 1993; Grakoui et al., 1993(a); Hijikata et al., 1993; Love et al., 1996; reviewed in Kwong et al., 1998). The C-terminal segment of the NS3 protein also harbors nucleoside triphosphatase and RNA helicase activity (Kim et al., 1995). The function of the protein NS4B is unknown. NS5A, a highly phosphorylated protein, seems to be responsible for the Interferon resistance of various HCV genotypes (Gale Jr. et al. 1997 Virology 230, 217; Reed et al., 1997. NS5B is an RNA-dependent RNA polymerase (RdRp) that is involved in the replication of HCV.

The open reading frame of the HCV RNA genome is flanked on its 5' end by a non-translated region (NTR) of approx. 340 nucleotides that functions as the internal ribosome entry site (IRES), and on its 3' end by a NTR of approximately 230 nucleotides. Both the 5' and 3' NTRs are important for RNA genome replication. The genomic sequence variance is not evenly distributed over the genome and the 5'NTR and parts of the 3'NTR are the most highly conserved portions. The authentic, highly conserved 3'NTR is the object of US patent 5,874,565 granted to Rice *et al.*

The cloned and characterized partial and complete sequences of the HCV genome have also been analyzed with regard to appropriate targets for a prospective antiviral therapy. Four viral enzyme activities provide possible targets such as (1) the NS2/3 protease; (2) the NS3/4A protease complex, (3) the NS3 Helicase and (4) the NS5B

3

RNA-dependent RNA polymerase. The NS3/4A protease complex and the NS3 helicase have already been crystallized and their three-dimensional structure determined (Kim et al., 1996; Yem et al., 1998; Love et al., 1996; Kim et al., 1998; Yao et al., 1997; Cho et al., 1998). The NS5B RNA dependent RNA polymerase has also been crystallized to reveal a structure reminiscent of other nucleic acid polymerases (Bressanelli et al. 1999, Proc. Natl. Acad. Sci, USA 96: 13034-13039; Ago et al. 1999, Structure 7: 1417-1426; Lesburg et al. 1999, Nat. Struct. Biol. 6: 937-943).

5

20

25

30

35

Even though important targets for the development of a therapy for chronic HCV infection have been defined with these enzymes and even though a worldwide intensive search for suitable inhibitors is ongoing with the aid of rational drug design and HTS, the development of therapy has one major deficiency, namely the lack of cell culture systems or simple animal models, which allow direct and reliable propagation of HCV viruses. The lack of an efficient cell culture system is still the main reason to date that an understanding of HCV replication remains elusive.

Although flavi- and pestivirus self-replicating RNAs have been described and used for the replication in different cell lines with a relatively high yield, similar experiments with HCV have not been successful to date (Khromykh *et al.*, 1997; Behrens *et al.*, 1998; Moser *et al.*, 1998). It is known from different publications that cell lines or primary cell cultures can be infected with high-titer patient serum containing HCV (Lanford *et al.* 1994; Shimizu *et al.* 1993; Mizutani *et al.* 1996; Ikda *et al.* 1998; Fourner *et al.* 1998; Ito *et al.* 1996). However, these virus-infected cell lines or cell cultures do not allow the direct detection of HCV-RNA or HCV antigens.

It is also known from the publications of Yoo et al. 1995; and of Dash et al., 1997; that hepatoma cell lines can be transfected with synthetic HCV-RNA obtained through in vitro transcription of the cloned HCV genome. In both publications the authors started from the basic idea that the viral HCV genome is a plus-strand RNA functioning directly as mRNA after being transfected into the cell, permitting the synthesis of viral proteins in the course of the translation process, and so new HCV particles could form HCV viruses and their RNA detected through RT-PCR. However the published results of the RT-PCR experiments indicate that the HCV replication in the described HCV transfected hepatoma cells is not particularly

10

efficient and not sufficient to measure the quality of replication, let alone measure the modulations in replication after exposure to potential antiviral drugs. Furthermore it is now known that the highly conserved 3' NTR is essential for the virus replication (Yanagi *et al.*, 1999). This knowledge strictly contradicts the statements of Yoo *et al.* (*supra*) and Dash *et al.* (*supra*), who used for their experiments only HCV genomes with shorter 3' NTRs and not the authentic 3' end of the HCV genome.

In WO 98/39031, Rice et al. disclosed authentic HCV genome RNA sequences, in particular containing: a) the highly conserved 5'-terminal sequence "GCCAGCC"; b) the HCV polyprotein coding region; and c) 3'-NTR authentic sequences.

In WO 99/04008, Purcell et al. disclosed an HCV infectious clone that also contained only the highly conserved 5'-terminal sequence "GCCAGC".

Recently Lohman *et al.* **1999** (Science 285: 110-113) and Bartenschlager *et al.* (in CA 2,303,526, laid-open on October 3, 2000) disclosed a HCV cell culture system where the viral RNA (I377/NS2-3') self-replicates in the transfected cells with such efficiency that the quality of replication can be measured with accuracy and reproducibility. The Lohman and Bartenschlager disclosures were the first demonstration of HCV RNA replication in cell culture that was substantiated through direct measurement by Northern blots. This replicon system and sequences disclosed therein highlight once again the conserved 5' sequence "GCCAGC". A similar observation highlighting the conservation of the 5'NTR was made by Blight *et al.* **2000** (Science 290: 1972-1974) and WO 01/89364 published on Nov. 29, 2001.

25

30

35

;

In addition to the conservation of the 5' and 3' untranslated regions in cell culture replicating RNAs, three other publications by Lohman *et al.* **2001**, Krieger *et al.* **2001** and Guo *et al.* **2001** have recently disclosed distinct adaptive mutants within the HCV non-structural protein coding region. Specific nucleotide changes that after the amino acids of the HCV non-structural proteins are shown to enhance the efficiency of establishing stable replicating HCV subgenomic replicons in culture cells.

Applicant has now found that, contrary to all previous reports, the highly conserved 5'-NTR can be mutated by adaptation to give rise to a HCV RNA sequence that, in conjunction with mutations in the HCV non-structural region, provides for a greater

efficiency of transduction and/or replication.

Applicant has also identified novel adaptive mutations within the HCV non-structural region that improves the efficiency of establishing persistently replicating HCV RNA in cell culture.

One advantage of the present invention is to provide an alternative to these existing systems comprising a HCV RNA molecule that self-replicates. Moreover, the present invention demonstrates that the initiating nucleotide of the plus-strand genome can be either an A as an alternative to the G already disclosed.

A further advantage of the present invention is to provide a unique HCV RNA molecule that transduces and/or replicates with higher efficiency. The Applicant demonstrates the utility of this specific RNA molecule in a cell line and its use in evaluating a specific inhibitor of HCV replication.

SUMMARY OF THE INVENTION

In a first embodiment, the present invention provides a 5'-non translated region of the hepatitis C virus wherein its highly conserved guanine at position 1 is substituted for adenine.

Particularly, the present invention provides a hepatitis C virus polynucleotide comprising adenine at position 1 as numbered according to the I377/NS2-3' construct (Lohmann et al. 1999, Accession # AJ242651).

Particularly, the invention provides a HCV self-replicating polynucleotide comprising a 5'-terminus consisting of ACCAGC (SEQ ID NO. 8).

In a second embodiment, the present invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.

5

10

15

25

Particularly, the invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as described above, further comprising the amino acid substitution E(1202)G.

More particularly, the invention provides a HCV self-replicating polynucleotide encoding a polyprotein comprising a G2042C or a G2042R mutation.

Most particularly, the invention provides for HCV self-replicating polynucleotide comprising a nucleotide substitution G—>A at position 1, and said polynucleotide encodes a polyprotein further comprising a G2042C or a G2042R mutation.

Particularly, the polynucleotide of the present invention can be in the form of RNA or DNA that can be transcribed to RNA.

In a third embodiment, the invention also provides for an expression vector comprising a DNA form of the above polynucleotide, operably linked with a promoter.

According to a fourth embodiment, there is provided a host cell transfected with the self-replicating polynucleotide or the vector as described above.

20

10

In a fifth embodiment, the present invention provides a RNA replication assay comprising the steps of:

- incubating the host cell as described above in the absence or presence of a potential hepatitis C virus inhibitor;
- 25 isolating the total cellular RNA from the cells;
 - analyzing the RNA so as to measure the amount of HCV RNA replicated;
 - comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor.
- In a sixth embodiment, the invention is directed to a method for testing a compound for inhibiting HCV replication, including the steps of:
 - a) treating the above described host cell with the compound:
 - b) evaluating the treated host cell for reduced replication, wherein reduced replication indicates the ability of the compound to inhibit replication.

WO 02/052015

5

10

7

DETAILED DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic view of the bi-cistronic replicon RNA. The sequence deviations between the I377/NS2-3' replicon from Lohman *et al.*, 1999 and the APGK12 replicon are indicated below the replicon. In place of a G nucleotide at the +1 position in the I377/NS2-3'replicon, the APGK12 contains an additional G resulting in GG at the 5' terminus (the first G being counted as position –1). In the linker region between the neo gene and the EMCV IRES sequence two areas deviate from I377/NS2-3': 14 nucleotides (CGCGCCCAGATGTT) which are not present in I377/NS2/3 are inserted at position 1184 in APGK12; 11 nucleotides (1231-1241) present in I377/NS2-3' are deleted to generate APGK-12. In the NS5B coding region, a T at position 8032 was mutated to C to eliminate a Ncol restriction site.

- 15 Figure 2 shows Northern blots of RNA-transfected Huh-7 cell lines. 12 µg of total . cellular RNA or control RNA was separated on 0.5% agarose-formaldehyde gels and transferred to Hybond N+ paper, fixed and (Figure 2A) radioactively probed with HCV specific minus-strand RNA that detects the presence of plus-strand replicon RNA. Lanes 1 and 2: positive controls that contain 109 copies of in vitro transcribed APGK12 RNA. Lane 3: negative control of total cellular RNA from untransfected 20 Huh-7 cells. Lanes 4 and 5: cellular RNA from B1 and B3 cell lines that have integrated DNA copies of the neomycin phosphotransferase gene. Lane 6: total cellular RNA from a Huh-7 cell line, designated S22.3, that harbors high copy number HCV sub-genomic replicon RNA as highlighted by the arrow. Other cell lines have no detectable replicon RNA. Figure 2B is identical to Figure 2A with the 25 exception that the blot was radioactively probed with HCV specific plus-strand RNA to detect the presence of HCV minus-strand RNA. Lanes 1 and 2 are positive control lanes that contain 109 copies of full length HCV minus strand RNA. Lane 6, which contains 12 µg of total cellular RNA from cell line S22.3, harbors detectable minusstrand replicon RNA at the expected size of 8 - 9 kilobases. M represent the 30 migration of non-radioactive molecular size markers on the agarose gel. 28s represents the migration of 28s ribosomal RNA and accounts for the detection of this species in a samples of total cellular RNA.
- Figure 3 shows indirect immunofluorescence of a HCV non-structural protein in the

8

S22.3 cell line. Indirect immunofluorescence was performed on cells that were cultured and fixed, permeabilized and exposed to a rabbit polyclonal antibody specific for a segment of the HCV NS4A protein. Secondary goat anti-rabbit antibody conjugated with red-fluor Alexa 594 (Molecular Probes) was used for detection. Top panels shows the results of immunofluorescence (40X objective) and the specific staining of the S22.3 cells. The bottom panels represent the identical field of cells viewed by diffractive interference contrast (DIC) microscopy. The majority of S22.3 (Figure 3A) cells within the field stain positively for HCV NS4A protein that localizes in the cytoplasm, whereas the B1 cells (Figure 3B) that fail to express any HCV proteins, only have background level of staining.

5

10

15

20

Figure 4 shows Western-blots following SDS-PAGE separation of total proteins extracted from three cell lines: (i) naïve Huh-7 cell line, (ii) neomycin resistant Huh-7 cell line B1, and (iii) the S22.3 cell line. Panels A, B, and C, demonstrate the results of western blots probed with rabbit polyclonal antisera specific for neomycin phosphotransferase (NPT), HCV NS3, and HCV NS5B, respectively. Visualization was achieved through autoradiographic detection of a chemiluminescent reactive secondary \ goat anti-rabbit antibody. Panel A shows that the S22.3 RNA replicon cell line, expresses the NPT protein at levels higher than control B1 cells and that the naïve Huh-7 cell line does not produce the NPT protein. Panels B and C show that only the S22.3 cell line produces the mature HCV NS3 and NS5B proteins, respectively. M represents molecular weight (in kilodaltons) of pre-stained polypeptide markers.

Figure 5A and 5B identify the nucleotide and amino acid sequences respectively that differ from the APGK12 sequence in the different HCV bi-cistronic replicons. The S22.3 adapted replicon is a first generation replicon selected following the transfection of RNA transcribed from the APGK12 template. R3, R7, R16 are second generation replicons that were selected following the transfection of RNA isolated from the S22.3 first generation replicon cell line. Figure 5A: Nucleotide mutations that were characterized in each of the adapted replicons are indicated adjacent to the respective segment of the replicon (IRES, NS3, NS4A, NS5A, and NS5B). Figure 5B: Amino acid numbers are numbered according to the full length HCV poly-protein with the first amino acid in the second cistron corresponding to amino acid 810 in NS2 of I377/NS2-3' construct.

9

Figure 6 depicts the colony formation efficiency of four in vitro transcribed HCV subgenomic bi-cistronic replicon RNAs. The APGK12 serves as the reference sequence; highlighted are the initiating nucleotides of the HCV IRES in each of the constructs and the amino acid differences (from the APGK12 reference sequence) in the HCV non-structural region for the two R3-rep. Note that the in vitro transcribed APGK-12 RNAs that harbor either a 5'G or 5'A form colonies with the same efficiency (ca. 80 cfu/µg in panels A and B) following selection with 0.25 mg/ml G418. RNA isolated from the second generation R3 cell line was reverse transcribed into DNA and cloned into the pAPGK12 vector backbone to generate the R3-rep. which was sequenced and found to encode additional changes that included the L(2155)P substitution in the NS5A segment of the HCV polyprotein (compare R3-rep sequence with the R3 sequence in tables 2 and 3). Various quantities of in vitro transcribed R3-rep-5'A RNA, were transfected into naïve Huh-7 cells to determine a colony formation efficiency of 1.2 X 10⁶ cfu/µg of RNA (panel C). Various quantities of R3-rep-5'G were also transfected resulting in a colony formation efficiency of 2 X 10⁶ cfu/μg of RNA (panel D).

Figure 7 displays a typical RT-PCR amplification plot (left panel) and the graphical representation of Ct values versus known HCV RNA quantity in a standard curve (right panel). Each of the plotted curves in the left panel, graph the increment of fluorescence reporter signal (delta-Rn) versus PCR cycle number for a predetermined quantity of HCV replicon RNA. The Ct value is obtained by determining the point at which the fluorescence exceeds an arbitrary value (horizontal line). The right panel demonstrates the linear relationship between starting RNA copy number of the predetermined standards (large black dots) and the Ct value. Smaller dots are the Ct values of RNA samples (containing unknown quantity of HCV replicon RNA) from S22.3 cells treated with various concentrations of a specific inhibitor of HCV replication.

30

35

5

10

15

20

25

Figure 8 shows the effect of increasing concentration of inhibitor A on HCV RNA replicon levels in Huh7 cells. S22.3 cells were grown in the presence of increasing concentrations of inhibitor A starting at 0.5nM and ranging to 1024nM. The inhibitor dose-response curve is the result of 11 concentrations from serial two-fold dilutions (1:1). One control well, without any inhibitor, was also included during the course of

the experiment. The cells were incubated for 4 days in a 5% CO₂ incubator at 37 °C. Total cellular RNA was extracted, quantified by optical density. HCV replicon RNA was evaluated by real time RT-PCR and plotted as genome equivalents/µg total RNA as a function of inhibitor concentration

5

10

15

Definitions

Unless defined otherwise, the scientific and technological terms and nomenclature used herein have the same meaning as commonly understood by a person of ordinary skill to which this invention pertains. Generally, the procedures for cell culture, infection, molecular biology methods and the like are common methods used in the art. Such standard techniques can be found in reference manuals such as for example Sambrook *et al.* (1989) and Ausubel *et al.* (1994).

Nucleotide sequences are presented herein by single strand, in the 5' to 3' direction, from left to right, using the one letter nucleotide symbols as commonly used in the art and in accordance with the recommendations of the IUPAC-IUB Biochemical Nomenclature Commission (1972).

The present description refers to a number of routinely used recombinant DNA (rDNA) technology terms. Nevertheless, definitions of selected examples of such rDNA terms are provided for clarity and consistency.

- The term "DNA segment or molecule or sequence", is used herein, to refer to molecules comprised of the deoxyribonucleotides adenine (A), guanine (G), thymine (T) and/or cytosine (C). These segments, molecules or sequences can be found in nature or synthetically derived. When read in accordance with the genetic code, these sequences can encode a linear stretch or sequence of amino acids which can be referred to as a polypeptide, protein, protein fragment and the like.
 - As used herein, the term "gene" is well known in the art and relates to a nucleic acid sequence defining a single protein or polypeptide. The polypeptide can be encoded by a full-length sequence or any portion of the coding sequence, so long as the functional activity of the protein is retained.
- A "structural gene" defines a DNA sequence which is transcribed into RNA and translated into a protein having a specific structural function that constitute the viral particles. "Structural proteins" defines the HCV proteins incorporated into the virus particles namely, core "C", E1, E2, and E2-p7.
- "Non-structural proteins", defines the HCV proteins that are not comprised in viral particles namely, NS2, NS3, NS4A, NS5A and NS5B.

11

"Restriction endonuclease or restriction enzyme" is an enzyme that has the capacity to recognize a specific base sequence (usually 4, 5 or 6 base pairs in length) in a DNA molecule, and to cleave the DNA molecule at every place where this sequence appears. An example of such an enzyme is *EcoRI*, which recognizes the base sequence G↓AATTC and cleaves a DNA molecule at this recognition site. "Restriction fragments" are DNA molecules produced by the digestion of DNA with a restriction endonuclease. Any given genome or DNA segment can be digested by a particular restriction endonuclease into at least two discrete molecules of restriction fragments.

5

25

- "Agarose gel electrophoresis" is an analytical method for fractionating polynucleotide molecules based on their size. The method is based on the fact that nucleic acid molecules migrate through a gel as through a sieve, whereby the smallest molecule has the greatest mobility and travels the farthest through the gel. The sieving characteristics of the gel retards the largest molecules such that, these have the
 least mobility. The fractionated polynucleotides can be visualized by staining the gel using methods well known in the art, nucleic acid hybridization or by tagging the fractionated molecules with a detectable label. All these methods are well known in the art, specific methods can be found in Ausubel et al. (supra).
 "Oligonucleotide or oligomer" is a molecule comprised of two or more
- deoxyribonucleotides or ribonucleotides, preferably more than three. The exact size of the molecule will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. An oligonucleotide can be derived synthetically, by cloning or by amplification.
 - "Sequence amplification" is a method for generating large amounts of a target sequence. In general, one or more amplification primers are annealed to a nucleic acid sequence. Using appropriate enzymes, sequences found adjacent to, or in between the primers are amplified. An amplification method used herein is the polymerase chain reaction (PCR) and can be used in conjunction with the reverse-transcriptase (RT) to produce amplified DNA copies of specific RNA sequences.
- 30 "Amplification primer" refers to an oligonucleotide, capable of annealing to a RNA or DNA region adjacent to a target sequence and serving as the initiation primer for DNA synthesis under suitable conditions well known in the art. The synthesized primer extension product is complementary to the target sequence.
- The term "domain" or "region" refers to a specific amino acid sequence that defines either a specific function or structure within a protein. As an example herein, is the

10

· 15

20

25

30

35

NS3 protease domain comprised within the HCV non-structural polyprotein. The terms "plasmid" "vector" or "DNA construct" are commonly known in the art and refer to any genetic element, including, but not limited to, plasmid DNA, phage DNA, viral DNA and the like which can incorporate the oligonucleotide sequences, or sequences of the present invention and serve as DNA vehicle into which DNA of the present invention can be cloned. Numerous types of vectors exist and are well known in the art.

The terminology "expression vector" defines a vector as described above but designed to enable the expression of an inserted sequence following transformation or transfection into a host. The cloned gene (inserted sequence) is usually placed under the control of control element sequences such as promoter sequences. Such expression control sequences will vary depending on whether the vector is designed to express the operably linked gene *in vitro* or *in vivo* in a prokaryotic or eukaryotic host or both (shuttle vectors) and can additionally contain transcriptional elements such as enhancer elements, termination sequences, tissue-specificity elements, and/or translational initiation and termination sites.

A host cell or indicator cell has been "transfected" by exogenous or heterologous DNA (e.g. a DNA construct) or RNA, when such nucleic acid has been introduced inside the cell. The transfecting DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transfecting/transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, an example of a stably transfected cell is one in which the transfecting DNA has become integrated into a chromosome and is inherited by daughter cells through chromosome replication. A host cell or indicator cell can be transfected with RNA. A cell can be stably transfected with RNA if the RNA replicates and copies of the RNA segregate to daughter cells upon cell division. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cells containing the transfecting DNA or RNA. Transfection methods are well known in the art (Sambrook et al., 1989; Ausubel et al., 1994). If the RNA encodes for a genetic marker that imparts an observable phenotype, such as antibiotic resistance, then the stable transfection of replicating RNA can be monitored by the acquisition of such phenotype by the host cell.

As used herein the term "transduction" refers to the transfer of a genetic marker to host cells by the stable transfection of a replicating RNA.

10

15

20

PCT/CA01/01843

The nucleotide sequences and polypeptides useful to practice the invention include without being limited thereto, mutants, homologs, subtypes, quasi-species, alleles, and the like. It is understood that generally, the sequences of the present invention encode a polyprotein. It will be clear to a person skilled in the art that the polyprotein of the present invention and any variant, derivative or fragment thereof, is auto-processed to an active protease.

As used herein, the designation "variant" denotes in the context of this invention a sequence whether a nucleic acid or amino acid, a molecule that retains a biological activity (either functional or structural) that is substantially similar to that of the original sequence. This variant may be from the same or different species and may be a natural variant or be prepared synthetically. Such variants include amino acid sequences having substitutions, deletions, or additions of one or more amino acids, provided the biological activity of the protein is conserved. The same applies to variants of nucleic acid sequences which can have substitutions, deletions, or additions of one or more nucleotides, provided that the biological activity of the sequence is generally maintained.

The term "derivative" is intended to include any of the above described variants when comprising additional chemical moiety not normally a part of these molecules. These chemical moieties can have varying purposes including, improving a molecule's solubility, absorption, biological half life, decreasing toxicity and eliminating or decreasing undesirable side effects. Furthermore, these moieties can be used for the purpose of labeling, binding, or they may be comprised in fusion product(s). Different moieties capable of mediating the above described effects can be found in *Remington's The Science and Practice of Pharmacy* (1995).

Methodologies for coupling such moieties to a molecule are well known in the art.

The term "fragment" refers to any segment of an identified DNA, RNA or amino acid sequence and/or any segment of any of the variants or derivatives described herein above that substantially retains its biological activity (functional or structural) as required by the present invention.

The terms "variant", "derivative", and "fragment" of the present invention refer herein to proteins or nucleic acid molecules which can be isolated/purified, synthesized chemically or produced through recombinant DNA technology. All these methods are well known in the art. As exemplified herein below, the nucleotide sequences and polypeptides used in the present invention can be modified, for example by *in vitro* mutagenesis.

20

As used herein, the term "HCV polyprotein coding region" means the portion of a hepatitis C virus that codes for the polyprotein open reading frame (ORF). This ORF may encode proteins that are the same or different than wild-type HCV proteins. The ORF may also encode only some of the functional protein encoded by wild-type polyprotein coding region. The protein encoded therein may also be from different isolates of HCV, and non-HCV protein may also be encoded therein.

As used herein, the abbreviation "NTR" used in the context of a polynucleotide molecule means a non-translated region. The term "UTR" means untranslated region. Both are used interchangeably.

Preferred embodiments

Particularly, the invention provides a HCV self-replicating polynucleotide molecule comprising a 5'-terminus consisting of ACCAGC (SEQ ID NO.8).

According to the first embodiment of this invention, there is particularly provided a HCV polynucleotide construct comprising:

- a 5'-non translated region (NTR) comprising the sequence ACCAGC at, or proximal to, its 5'-terminus;
- a HCV polyprotein coding region; and
- a 3'-NTR region.

In a second embodiment, the present invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.

Particularly, the invention is directed to a HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as described above, further comprising the amino acid substitution E(1202)G.

Alternatively, the first embodiment of the present invention is directed to HCV selfreplicating polynucleotide molecule comprising a G2042C/R mutation.

20

25

30

35

According to the second embodiment, the present invention particularly provides a HCV polynucleotide construct comprising:

- a 5'-NTR region comprising the sequence ACCAGC at, or proximal to, its 5'-terminus;
- a HCV polyprotein region coding for a HCV polyprotein comprising a
 G(2042)C or a G(2042)R mutation; and
- a 3'-NTR region.
- 10 Preferably, the polynucleotide construct of the present invention is a DNA or RNA molecule. More preferably, the construct is a RNA molecule. Most preferably, the construct is a DNA molecule.
- More particularly, the first embodiment of this invention is directed to a RNA

 molecule encoded by the DNA molecule selected from the group consisting of: SEQ

 ID NO. 2, 4, 5, 6, 7, 24 and 25.
 - Most particularly, the invention provides a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.

In a third embodiment, the invention also is directed to an expression vector comprising DNA forms of the above polynucleotide, operably linked with a promoter.

Preferably, the promoter is selected from the group consisting of: T3, T7 and SP6.

According to a fourth embodiment, there is provided a host cell transfected with the self-replicating polynucleotide or vector as described above. Particularly, the host cell is a eukaryotic cell line. More particularly, the eukaryotic cell line is a hepatic cell line. Most particularly, the hepatic cell line is Huh-7.

In a fifth embodiment, the present invention provides a RNA replication assay comprising the steps of:

- a) incubating the host cell as described above under conditions suitable for RNA replication;
- b) isolating the total cellular RNA from the cells; and

c) analyzing the RNA so as to measure the amount of HCV RNA replicated.

Preferably, the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.

Alternatively in this fifth embodiment, the construct comprises a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.

10

15

5

According to a preferred aspect of the sixth embodiment, the invention is directed to a method for testing a compound for inhibiting HCV replication, including the steps of:

- a) carrying step a) as described in the above assay, in the presence or absence of the compound;
- b) isolating the total cellular RNA from the cells; and
- c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
- d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor,
- wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

Preferably, the cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.

25

EXAMPLES

EXAMPLE 1

Replicon Constructs (APGK-12; Figure 1)

pET9a-EMCV was obtained by ligating an oligonucleotide linker
5' gaattccagatggcgcccagatgttaaccagatccatggcacactctagagtactgtcgac 3' (SEQ ID NO.9) to pET-9a (Novagen) that was cut with EcoRI and SalI to form the vector pET-9a-mod. This linker contains the following restriction sites: EcoRI, AscI, HpaI, NcoI, XbaI, ScaI, SalI. The EMCV IRES was amplified by PCR from the vector pTM1 with primers

17

5' cggaatcgttaacagaccacaacggtttccctc 3' (SEQ ID NO.10) and 5' ggcgtacccatggtattatcgtgtttttca 3' (SEQ ID NO.11) and ligated into pET-9a-mod via EcoRI and Ncol to form pET-9a-EMCV.

- The sequence of HCV NS2 to NS5B followed by the 3'UTR of HCV was obtained from the replicon construct I377/NS2-3' (Lohman *et al.*, 1999; accession number: AJ242651) and synthesized by Operon Technologies Inc. with a T to C change at the Ncol site in NS5B at nucleotide 8032. This sequence was released from an GenOp® vector (Operon Technologies) with Ncol and Scal and transferred into pET-9a-EMCV-NS2-5B-3'UTR.
 - pET-9a-HCV-neo was obtained by amplification of the HCV IRES from a HCV cDNA isolated from patient serum with primers
 - 5' gcatatgaattctaatacgactcactataggccagcccccgattg 3' (SEQ ID NO.12) containing a T7 promoter and primer
 - 5' ggcgcgccctttggtttttctttgaggtttaggattcgtgctcat 3' (SEQ ID NO.13) and amplification of the neomycin phosphotransferase gene from the vector pcDNA 3.1 (Invitrogen) with primers
- 5' aaagggcgcatgattgaacaagatggattgcacgca 3' (SEQ ID NO.14) and 5'
 gcatatgttaactcagaagaactcgtcaagaaggcgata 3' (SEQ ID NO.15). These two PCR fragments were mixed and amplified with primers
 - 5' gcatatgaattctaatacgactcactataggccagccccgattg 3' (SEQ ID NO.16) and 5' gcatatgttaactcagaagaactcgtcaagaaggcgata 3' (SEQ ID NO.15), cut with Eco RI and HpaI and transferred into pET-9a-mod to form pet-9a-HCV-neo. The EMCV-
- NS2-5B-3'UTR was released from pET-9a-EMCV-NS2-5B-3'UTR with HpaI and ScaI and transferred into pet-9a-HCV-neo that was cut with HpaI to form pET-9a-APGK12. This insert was sequenced with specific successive primers using a ABI Prism® BigDye™ Terminator Cycle sequencing kit and analyzed on ABI Prism® 377 DNA Sequencer and is shown in SEQ ID NO 1.

30

35

15

RNA in vitro transcription

pET-9a-APGK12 DNA was cut with Scal for expression of the full-length replicon or with BgllI for expression of a truncated negative control RNA. DNA was analyzed on a 1% agarose gel and purified by Phenol/Chloroform extraction. RNA was produced using a T7 Ribomax® kit (Promega) followed by extraction with phenol/chloroform

and precipitation with 7.5 M LiCl₂. RNA was treated with DNAse I for 15 min to remove the DNA template and further purified with an RNeasy® column (Qiagen). RNA integrity was verified on a denaturing formaldehyde 1% agarose gel.

5 EXAMPLE 2

20

25

30

Primary transfection of Huh? cells and selection of replicon cell lines

Human hepatoma Huh? cells (Health Science Research Resources Bank, Osaka,
Japan) were grown in 10% FBS/DMEM. Cells were grown to 70% confluency,
trypsinized, washed with phosphate buffered saline (PBS) and adjusted to 1x10⁷

cells/ml of PBS. 800 μl of cells were transferred into 0.4cm cuvettes and mixed with
15 μg of replicon RNA. Cells were electroporated using 960μF, 300 volts for ~18
msec and evenly distributed into two 15 cm tissue culture plates and incubated in a
tissue culture incubator for 24 hours. The selection of first and second generation
replicon cell lines was with 10% FBS/DMEM medium supplemented with 1mg/ml of
G418. Cells were selected for 3-5 weeks until colonies were observed that were
isolated and expanded.

Following the G418 selection and propagation of Huh-7 cells transfected with APGK12 (SEQ ID NO. 1) RNA, cells that formed a distinct colony were treated with trypsin and serially passed into larger culture flasks to establish cell lines. Approximately 10 X 10⁶ cells were harvested from each cell line. The cells were lysed and the total cellular RNA extracted and purified as outlined in Qiagen RNAeasy® preparatory procedures. Figure 2 shows the analysis of 12 µg of total cellular RNA from various cell lines as analyzed on a Northern blot of a denaturing agarose-formaldehyde gel.

Figure 2A is a Northern blot (radioactively probed with HCV specific minus-strand RNA) that detects the presence of plus-strand replicon RNA. Lanes 1 and 2 are positive controls that contain 10⁹ copies of in vitro transcribed APGK12 RNA. Lane 2 contains the *in vitro* transcribed RNA mixed with 12 µg of total cellular from naïve Huh-7 cells. Lane 3 is a negative control of total cellular RNA from untreated Huh-7 cells. Lanes 4 and 5 contain cellular RNA from the B1 and B3 G418 resistant cell lines that have DNA integrated copies of the neomycin phosphotransferase gene. Lane 6 contains total cellular RNA from a Huh-7 cell line, designated S22.3, that

19

harbors high copy number of HCV sub-genomic replicon RNA as detected by the positive signal in the 8 kilo-base range. Other cell lines have no detectable replicon RNA. Figure 2B is a Northern blot of a duplicate of the gel presented in 2A with the exception that the blot was radioactively probed with HCV specific plus-strand RNA to detect the presence of HCV minus-strand RNA (lanes 1 and 2 are positive control lanes that contain 10^9 copies of full length genomic HCV minus strand RNA); only lane 6, which contains $12~\mu g$ of total cellular RNA from cell line S22.3, harbors detectable minus-strand replicon RNA at the expected size of 8-9 kilobases. An quantitative estimation of RNA copy number, based on phosphorimager scanning of the Northern blots, is approximately $6~X10^7$ copies of plus-strand/ μg of total RNA, and $6~x~10^6$ copies of minus strand/ μg of total RNA. The presence of the plus-strand and minus-strand intermediate confirms that the HCV sub-genomic RNA is actively replicating in the S22.3 cell line.

15 EXAMPLE 3

5

10

20

25

30

35

S22.3 cell line constitutively expresses HCV non-structural proteins.

HCV non-structural protein expression was examined in the S22.3 cell line. Figure 3 displays the result of indirect immunofluorescence that detects the HCV NS4A protein in the S22.3 cell line and not in the replicon negative B1 cell line (a G418 resistant Huh-7 cell line). Indirect immunofluorescence was performed on cells that were cultured and fixed (with 4% paraformaldehyde) onto Lab-tek chamber slides. Cells were permeabilized with 0.2% Triton X-100 for 10 minutes followed by a 1 hour treatment with 5% milk powder dissolved in phosphate-buffered saline (PBS). A rabbit serum containing polyclonal antibody raised against a peptide spanning the HCV NS4A region was the primary antibody used in detection. Following a 2 hour incubation with the primary antibody, cells were washed with PBS and a secondary goat anti-rabbit antibody conjugated with red-fluor Alexa® 594 (Molecular Probes) was added to cells for 3 hours. Unbound secondary antibody was removed with PBS washes and cells were sealed with a cover slip. Figure 3 (top panels) shows the results of immunofluorescence as detected by a microscope with specific fluorescent filtering; the bottom panels represent the identical field of cells viewed by diffractive interference contrast (DIC) microscopy. The majority of S22.3 (Figure 3A) cells within the field stain positively for HCV NS4A protein that localizes in the cytoplasm, whereas the B1 cells (Figure 3B) that fail to express any HCV proteins, only have

background level of staining. A small proportion of S22.3 cells express high levels of intensely stained HCV NS4A.

Expression of the proteins encoded by the bi-cistronic replicon RNA was also examined on Western-blots following SDS-PAGE separation of total proteins 5 extracted from: (i) naïve Huh-7 cell line, (ii) neomycin resistant Huh-7 cell line B1, and (iii) the S22.3 cell line. Figure 4 panels A, B, and C, demonstrate the results of western blots probed with rabbit polyclonal antisera specific for neomycin phosphotransferase (NPT), HCV NS3, and HCV NS5B, respectively. Visualization was achieved through autoradiographic detection of a chemiluminescent reactive 10 secondary HRP-conjugated goat anti-rabbit antibody. Figure 4 panel A shows that the S22.3 RNA replicon cell line, expresses the NPT protein at levels higher than B1 cells (which contain an integrated DNA copy of the npt gene) and that the naïve Huh-7 cell line does not produce the NPT protein. Figure 4 panels B and C show that only the S22.3 cell line produces the mature HCV NS3 and NS5B proteins, 15 respectively. The western blots demonstrate that the S22.3 cell line, which harbors actively replicating HCV sub-genomic replicon RNA, maintains replication of the RNA through the high level expression of the HCV non-structural proteins.

20 EXAMPLE 4

Sequence determination of adapted replicons

Total RNA was extracted from replicon containing Huh7 cells using a RNeasy Kit (Qiagen). Replicon RNA was reverse transcribed and amplified by PCR using a 25 OneStep RT-PCR kit (Qiagen) and HCV specific primers (as selected from the full-length sequence disclosed in WO 00/66623). Ten distinct RT-PCR products, that covered the entire bi-clstronic replicon in a staggered fashion, were amplified using oligonucleotide primers. The PCR fragments were sequenced directly with ABI Prism® BigDye™ Terminator Cycle PCR Sequencing and analyzed on ABI Prism® 377 DNA Sequencer. To analyze the sequence of the HCV replicon 3' and 5' ends a RNA ligation/RT-PCR procedure described in Kolykhalov *et al.* 1996 was followed. The nucleotide sequence of S22.3 is presented as SEQ ID NO. 2.

21

EXAMPLE 5.

5

10

15

20

25

30

Serial Passage of HCV Replicon RNA

The total cellular RNA from the S22.3 cell line was prepared as described above. HCV Replicon RNA copy number was determined by Taqman® RT-PCR analysis and 20 µg of total S22.3 cellular RNA (containing 1 X 109 copies of HCV RNA) was transfected by electroporation into 8 X 10⁶ naïve Huh-7 cells. Transfected cells were subsequently cultured in 10 cm tissue culture plates containing DMEM supplemented with 10% fetal calf serum (10% FCS). Media was changed to DMEM (10% FCS) supplemented with 1 mg/ml G418 24 hours after transfection and then changed every three days. Twenty-three visible colonies formed three to four weeks post-transfection and G418 selection. G418 resistant colonies were expanded into second generation cell lines that represent the first cell lines harboring serially passaged HCV Replicon RNA. Three of these cell lines: R3, R7, and R16 were the subject of further analyses. First, the efficiency of transduction by each of the adapted replicons was determined by electroporation of the total cellular RNA (extracted from the R3, R7 and R16) into naïve Huh-7 cells; following electroporation, the transduction efficiency was determined as described above, by counting the visible G418 resistant colonies that arose following 3 to 5 weeks of G418 selection (Table 1). Second, the sequence of the serially passed adapted replicons was determined from the total cellular RNA that was extracted from each of the R3, R7 and R16 replicon cell lines as described in example 4 (SEQ ID NO. 4, 5, 6). Using the pAPGK12 as a reference sequence (SEQ ID NO. 1), the nucleotide changes that were selected in HCV segment of the adapted replicons are presented in Figure 5A. Some of these nucleotide changes are silent and do not change the encoded amino acid whereas others result in an amino acid substitution. Figure 5B summarizes the amino acid changes encoded by the adapted replicons with the amino acid sequence of pAPGK12 as the reference. It is important to note that the reference sequence APGK-12 (SEQ ID NO.1) contains an extra G at the 5'-terminal (5'-GG) that is not maintained in the replicating RNA of the established cell lines. Also noteworthy is that, in addition to G->A at nucleotide 1, there is also an adapted mutation G->C/R at amino acid 2042 (shown as amino acid 1233 in the sequence listing since a.a. 810 of NS2 is numbered as a.a. 1 in SEQ ID) that can be found in all clones analyzed.

22

TABLE 1
Transfection of Huh-7 cells

	RNA	Copies of Replicon	# Colonies	SEQ ID
5				
	5 ng APKG12 replicon in 20μg total Huh-7 RNA	1.2 x 10 ⁹	0	
10	15 μg APKG12 replicon RNA	3 x 10 ¹²	1 (S22.3)	1
	20μg total:	3 x 10 ⁹	23 (3 clones	2
			analyzed)	-
15	R3 cellular RNA	1 x 10 ⁹	200	4
	R7 cellular RNA	1 x 10 ⁹	20	5
	R16 cellular RNA	3 x 10 ⁸	100	6
	cloned R3rep RNA	2.3 x 10 ⁸	2000	7

20 EXAMPLE 6

25

30

Construction of APGK12 with 5' G-> A substitution (APGK12-5'A, SEQ ID NO.24)

The pAPGK12 DNA was modified to change the first nucleotide in the sequence to replace the 5'GG with a 5'A. The change in the pAPGK12 was introduced by replacing an *EcoRI/AgeI* portion of the sequence with a PCR-generated *EcoRI/AgeI* fragment that includes the mutation. The oligonucleotides used for the amplification were (SEQ ID. NO. 20): 5'-GTG GAC GAA TTC TAA TAC GAC TCA CTA TAA CCA GCC CCC GAT TGG-3' and (SEQ ID. NO. 21): 5'-GGA ACG CCC GTC GTG GCC AGC CAC GAT-3' and generated a 195 bp DNA fragment that was then digested with *EcoRI* and *AgeI*. The resulting 178 bp restriction fragment was used to replace the *EcoRI / AgeI* fragment in pAPGK12 to generate the pAPGK12-5'A plasmid.

EXAMPLE 7

cDNA cloning of the R3-replicon (R3Rep).

The cDNA clone of the R3 replicon was produced by RT-PCR of RNA extracted from the R3 cell line. The following two oligonucleotides were used: (SEQ ID. NO. 22): 5'-GTC GTC TCT GAC ATG GAG AC-3' and (SEQ ID. NO. 23): 5'-GAG TTG

10

20

25

30

35

CTC AGT GGA TTG ATG GGC AGC-3'. The ~4400nt PCR fragment, starting within the NS2 coding region and extending to the 5'-end of the NS5B coding region, was cloned into the plasmid pCR3.1 by TA cloning (Invitrogen). The SacII / XhoI portion of this R3 sequence was then used to replace the SacII / XhoI fragment present in the pAPGK12 and the pAPGK12-5'A described above. Consequently, two R3 cDNA sequences were generated: (I) R3-Rep-5'G with an initiating 5'G (SEQ ID NO.7), and R3-Rep-5'A (SEQ ID NO.25) with an initiating 5'A. Sequencing of the R3 rep cDNA identified unique nucleotide changes that differ from the original pAPGK12 sequence (see Figure 5A); some of these changes are silent and do not change the encoded amino acid, whereas others do result in an amino acid change (see Figure 5B). The differences between R3 and the R3-rep reflect the isolation of a unique R3-rep cDNA clone encoding nucleotide changes that were not observed from the sequencing of the total RNA extracted from the R3 cell line.

15 EXAMPLE 8

Efficiency of colony formation with modified constructs

RNA from pAPGK12, pAPGK12-5'A, pR3-Rep and pR3-Rep-5'A was generated by in vitro transcription using the T7 Ribomax® kit (Promega) as described in example 1 above. The reactions containing the pAPGK12-5'A and pR3-Rep-5'A templates were scaled-up 10-fold due to the limitation of commercial RNA polymerase in initiating transcripts with 5'-A. The full length RNAs and control truncated RNA for each clone were introduced into 8 x 10⁶ naïve Huh-7 cells by electroporation as described in example 2. Replicon RNA was supplemented with total cellular Huh-7 carrier RNA to achieve a final 15-20µg quantity. The cells were then cultured in DMEM medium supplemented with 10% fetal calf serum and 0.25 mg/ml G418 in two 150 mm plates. The lower concentration of G418 was sufficient to isolate and select replicon containing cell lines as none of the transfectants with the control truncated RNA produced any resistant colonies. In contrast, in vitro transcribed APGK-12 RNAs that harbor either a 5'G or 5'A form colonies with the same efficiency (ca. 80 cfu/µg in Figure 6 panels A and B) following selection with G418. Various quantities (ranging from 0.1 ng to 1 µg) of the R3-rep-5'A RNA, were transfected into naïve Huh-7 cells to determine a colony formation efficiency of 1.2 X 10⁶ cfu/μg of RNA (Figure 6 panel C depicts transfection with 1 μg of RNA). Various quantities (ranging from 0.1 ng to 1 µg) of R3-rep [5'G] were similarly transfected resulting in a colony formation efficiency of 2 X 10⁶ cfu/µg of RNA (Figure 6 panel D

24

depicts colony formation with 1µg of RNA). Note that, shown for the first time, HCV subgenomic replicons replicate as efficiently with a 5' A nucleotide in place of the 5'G. APGK12 with a 5'A or 5'G RNA have similar transduction efficiencies. Similarly, R3-Rep RNAs with either the 5'A or 5'G both display the markedly increased transduction efficiency. Notably, the adaptive mutants within the HCV non-structural segment encoded by the R3-Rep provides for a substantial increase in transduction efficiency as depicted by the dramatic increase in colony forming units per µg of transfected RNA.

10 EXAMPLE 9

5

15

20

30

Quantification of HCV Replicon RNA Levels in Cell lines

S22.3 cells, or cell lines harboring other adapted replicons, were seeded in DMEM supplemented with 10% FBS, PenStrep and 1µg/mL Geneticin. At the end of the incubation period the replicon copy number is evaluated by real-time RT-PCR with the ABI Prism 7700 Sequence Detection System. The TAQMAN® EZ RT-PCR kit provides a system for the detection and analysis of HCV RNA (as first demonstrated by Martell *et al.* 1999 J. Clin. Microbiol. 37: 327-332). Direct detection of the reverse transcription polymerase chain reaction (RT-PCR) product with no downstream processing is accomplished by monitoring the increase in fluorescence of a dyelabeled DNA probe (Figure 6). The nucleotide sequence of both primers (adapted from Ruster, B. Zeuzem, S. and Roth, W.K., 1995. Analytical Biochemistry 224:597-600) and probe (adapted from Hohne, M., Roeske, H. and Schreier, E. 1998, Poster Presentation: P297 at the Fifth International Meeting on Hepatitis C Virus and Related Viruses Molecular Virology and Pathogenesis, Venezia-Lido Italy, June 25-28, 1998) located in the 5'-region of the HCV genome are the following:

HCV Forward primer:

5' ACG CAG AAA GCG TCT AGC CAT GGC GTT AGT 3' (SEQ ID NO.17)

HCV Reverse primer:

5' TCC CGG GGC ACT CGC AAG CAC CCT ATC AGG 3' (SEQ ID NO.18)

HCV Probe:

5' FAM-TGG TCT GCG GAA CGG GTG AGT ACA CC-TAMRA 3' (SEQ ID NO.19)

5 FAM: Fluorescence reporter dye.

TAMRA: Quencher dye.

Using The TAQMAN® EZ RT-PCR kit, the following reaction was set up:

Component	Volume per sample	Final	
	(µL)	Concentration	
RNase-Free Water	16	-	
5X Taqman EZ Buffer	10	1X	
Manganese Acetate 25mM	6	3mM	
dATP 10mM	1.5	300µM	
dCTP 10mM	1.5	300µM	
dGTP 10mM	1.5	300µM	
dUTP 20mM	1.5	300µM	
HCV Forward Primer 10µM	1	200nM	
HCV Reverse Primer 10µM	1	200nM	
HCV Probe 5uM	2	200nM	
rTth DNA Polymerase	2	. 0.1U/µL	
2.5U/µL			
AmpErase UNG 1U/μL	0.5	0.01U/µL	
Total Mix	45	-	

10

15

To this reaction mix, 5μ L of total RNA extracted from S22.3 cells diluted at $10 \text{ng}/\mu$ L was added, for a total of 50 ng of RNA per reaction. The replicon copy number was evaluated with a standard curve made from known amounts of replicon copies (supplemented with 50 ng of wild type Huh-7 RNA) and assayed in an identical reaction mix (Figure 7).

Thermal cycler parameters used for the RT-PCR reaction on the ABI Prism 7700 Sequence Detection System were optimized for HCV detection:

26

Cycle	Temperature (°C)	Time (Minutes)	Repeat	Reaction
Hold	50	2		Initial Step
Hold	60	30		Reverse
		•		Transcription
Hold	95	5		UNG Deactivation
L.VCIE	95 -	0:15	2	Melt
	60	1	2	Anneal/Extend
Cycle	90	0:15	40	Melt
	60	1	40	Anneal/Extend

Quantification is based on the threshold cycle, where the amplification plot crosses a defined fluorescence threshold. Comparison of the threshold cycles provides a highly sensitive measure of relative template concentration in different samples. Monitoring during early cycles, when PCR fidelity is at its highest, provides precise data for accurate quantification. The relative template concentration can be converted to RNA copy numbers by employing a standard curve of HCV RNA with known copy number (Figure 7).

10 **EXAMPLE 10**

15

20

25

A specific HCV NS3 protease anti-viral compound inhibits replication of the HCV replicon in S22.3 cell lines.

In order to determine the effect of a specific HCV NS3 protease anti-viral compound on replicon levels in S22.3 cells, the cells were seeded in 24 Well Cell Culture Cluster at 5 X 10⁴ cells per well in 500μL of DMEM complemented with 10% FBS, PenStrep and 1μg/mL Geneticin. Cells were incubated until compound addition in a 5% CO₂ incubator at 37 °C. The dose-response curve of the inhibitor displayed 11 concentrations resulting from serial two-fold dilutions (1:1). The starting concentration of compound A was 100nM. One control well (without any compound) was also included in the course of the experiment. The 24 well plates were incubated for 4 days in a 5% CO₂ incubator at 37 °C. Following a 4 day incubation period, the cells were washed once with PBS and RNA was extracted with the RNeasy® Mini Kit and Qiashredder® from Qiagen. RNA from each well was eluted in 50uL of H₂O. The RNA was quantified by optical density at 260nm on a Cary 1E UV-Visible Spectrophotometer. 50 ng of RNA from each well was used to quantify the HCV replicon RNA copy number as detailed in Example 6. The level of inhibition (% inhibition) of each well containing inhibitor was calculated with the following

equation (CN = HCV Replicon copy number):

$$\% \cdot inhibition = \left(\frac{CN \cdot control - CN \cdot well}{CN \cdot control}\right) * 100$$

The calculated % inhibition values were then used to determine IC_{50} , slope factor (n) and maximum inhibition (I_{max}) by the non-linear regression routine NLIN procedure of SAS using the following equation:

$$\% \cdot inhibition = \frac{I_{\text{max}} \times [inhibitor]^n}{[inhibitor]^n + IC_{50}^n}$$

10

Compound A was tested in the assay at least 4 times. The IC₅₀ curves were analyzed individually by the SAS nonlinear regression analysis. Figure 8 shows a typical curve and Table 2 shows the individual and average IC₅₀ values of compound A. The average IC₅₀ of compound A in the replication assay was 1.1nM.

15

TABLE 2

IC₅₀ of compound A in the S22.3 Cell line Replicon Assay.

Compound	IC ₅₀ (nM)	Average IC ₅₀ (nM)
	1.2	
Α	1.2	•
	1.0	
	0.9	
		1.1 + 0.2

20 DISCUSSION

25

The reproducible and robust *ex vivo* propagation of hepatitis C virus, to levels required for the accurate testing of potential anti-viral compounds, has not been achieved with any system. As an alternative approach to studying the molecular mechanisms of hepatitis C virus RNA replication, selectable self-replicating bicistronic RNAs were developed (Lohman *et al.*, 1999, Science 285:110-113; Bartenschlager CA 2,303,526). Minimally, these replicans encode for some or all of

28

the non-structural proteins and also carry a selectable marker such as the neomycin phosphotransferase. Though intracellular steady-state levels of these sub-genomic replicon RNAs among the selected clones is moderate to high, the frequency of generating G418-resistant colonies upon transfection of the consensus RNA described by Lohman et al. or Bartenschlager is very low. Less than 100 colonies are generated when 8 million cells are transfected with 1 µg of in vitro transcribed bicistronic replicon RNA. A low efficiency of colony formation was first noted by Lohmann et al (1999 et al, Science 285:110-113). Since then, Lohmann et al. (2001), Blight et al. (2000), and Guo et al. (2001), have isolated sub-genomic RNAs with markedly improved efficiencies in the colony formation assay. Lohmann et al., 1999 originally reported that selection of sub genomic replicons may not involve the selection of adaptive mutants as serially passaged RNA did not demonstrate an improved transfection efficiency. Nevertheless, in an effort to characterize the function and fitness of replicating HCV RNA, we serially passaged the replicon RNA that was isolated from the first selected cell-line. Notably, a significant increase in colony forming efficiency was obtained from this experiment, even though the quantity of replicon RNA was orders of magnitude lower than originally used to transfect the in vitro transcribed RNA. Furthermore, a second round serial passage of replicon RNA from this first generation clone into naive Huh-7 cells provided for yet another increase in colony formation efficiency (Table 1).

Our analysis of replicating HCV RNAs identified several adaptive mutations that enhance the efficiency of colony formation by up to 4 orders of magnitude. Adaptive mutations were found in many non-structural proteins, as well as in the 5' non-translated region. The substitution of the 5'-GG doublet for a 5'-A as the inaugurating nucleotide of the HCV 5'-UTR is a variant of the HCV genome that has not been previously described, despite the sequencing of innumerable genotypes and subtypes from across the world. Our original replicon that carried a 5'-GG evolved to variants with either a single 5'-A or 5'-G, both of which showed equal transduction efficiency. We describe here the first report of a HCV genome that can tolerate and stably maintain a 5'A extremity. Moreover, we were successful in re-introducing this defined single nucleotide substitution into our cDNA clone and generate *in vitro* transcribed RNA harboring such an extremity to confirm that a 5'A functions as efficiently as a 5'G.

5

10

15

20

25

30

We have identified adaptive amino acid substitutions in the HCV non-structural proteins NS3, NS4A and NS5A in the R3 replicon, and a substitution in NS5B in the R7 clone (see Figure 5B). These mutations, particularly the combination defined by the R3-rep (SEQ ID NO. 7), when reconstituted into a cDNA clone and transcribed onto a RNA replicon, result in a significantly enhanced transduction efficiency of up to 20,000 fold from the original wild type APGK12 replicon RNA. However, the steady state levels of intracellular replicon RNA were comparable from each of the different isolated clones. This result suggests that the increase in replication efficiency by the adaptive mutations does not result in higher stable intracellular RNA levels due to higher RNA replication, but rather confers increased permissivity for establishing the replicon in a greater number of Huh7 cells. Such a phenotype may be manifested transiently, through an initial increase of the amount of *de novo* replication, that is required to surpass a defined threshold to establish persistently replicating RNAs within a population of dividing cells.

Recently three other groups also identified other distinct adaptive mutants. Lohmann et al. (2000) reported enhanced transduction efficiencies of up to 10,000 fold with mutations in NS3, NS4B, NS5A and NS5B. Blight et al. (2000) reported an augmentation of transduction efficiencies up to 20,000 fold with a single mutation in NS5A whereas Guo et al. (2001) reported increases in transduction efficiencies of 5,000-10,000 fold with a deletion of a single amino acid in NS5A. The amino acid substitutions that we describe here have not previously been identified as adaptive mutants that enhance the efficiency of RNA transfection and/or replication. One exception is the mutation of E1202G in NS3 that we found in both the R7 and R16 replicons. This adaptation was previously described by Guo et al (2001) and Krieger et al (2001). All other adaptive mutations, without exception, described herein are unpublished.

The development of selectable subgenomic HCV replicons has provided for potential avenues of exploration on HCV RNA replication, persistence, and pathogenesis in cultured cells. However, the low transduction efficiency with the HCV RNA-containing replicons as originally described (Lohmann et al., 1999) showed that it was not a practical system for reverse genetics studies. The adaptive mutants described herein overcome the low transduction efficiency. In light of the recent descriptions of adaptive mutants by other groups, we note that adaptation can be

achieved by distinct mutations in different HCV NS proteins, although the level of adaptation can vary drastically. The replicons encoding adaptive mutants that are described herein are ideally suited for reverse genetic studies to identify novel HCV targets or host cell targets that may modulate HCV RNA replication or HCV replicon RNA colony formation. The adapted and highly efficient replicons are suitable tools for characterizing subtle genotypic or phenotypic changes that affect an easily quantifiable transduction efficiency.

Lastly, we have used our adapted HCV sub genomic replicon cell-line to

demonstrate the proficient inhibition of HCV RNA replication by a specific small molecule inhibitor of the HCV NS3 protease. This is the first demonstration that an antiviral, designed to specifically inhibit one of the HCV non-structural proteins, inhibits HCV RNA replication in cell culture. Moreover, this compound and our S22.3 cell line validate the proposal that RNA replication is directed by the HCV non-structural proteins NS3 to NS5B. The assay that we have described and validated will be extremely useful in characterizing other inhibitors of HCV non-structural protein function in cell culture in a high throughput fashion.

All references found throughout the present disclosure are herein incorporated by reference whether they be found in the following list or not.

References

Ago et al. 1999, Structure 7: 1417-1426

Ausubel et al., 1994, Current Protocols in Molecular Biology, Wiley, New York.

25 Bartenschlager, R. et al., 1993, J. Virol., 67, 3835-3844.

Behrens et al., 1998, J. Virol. 72, 2364

Bressanelli et al. 1999, Proc. Natl. Acad. Sci, USA 96: 13034-13039

Blight et al. 2000, Science 290: 1972-1974

Cho et al., 1998, J. Biol. Chem., 273, 15045

30 Dash *et al.*, **1997**, Am. J. Pathol., 151, 363 – 373

Fourner et al. 1998, J. Gen. Virol. 79, 2376

Gale Jr. et al. 1997 Virology 230, 217

Grakoui, A. et al., 1993(a), J. Virol. 67, 1385-1395.

Grakoui A, et al., 1993(b), Proc Natl Acad Sci USA, 90, 10583-7

35 Guo et al. (2001) J. Virol. 8516-8523

Hijikata, M. et al., 1991, Proc. Natl. Acad. Sci. USA. 88, 5547-5551.

Hijikata, M. et al., 1993, J. Virol. 67, 4665-4675.

Ikda et al. 1998, Virus Res. 56, 157

Ito et al. 1996, J. Gen. Virol. 77, 1043 - 1054

5 IUPAC-IUB Biochemical Nomenclature Commission, **1972**, Biochemistry, *11*, 1726-1732.

Kolykhalov et al. 1996 J. of Virology, 7, p. 3363-3371

Khromykh et al., 1997, J. Virol. 71, 1497

Kim, D.W. et al., 1995, Biochem. Biophys. Res. Comm., 215, 160-166.

10 Kim et al., 1996, Cell, 87, 343;

Kim et al., 1998, Structure, 6, 89

Kim et al., 1999, Arch. Virol, 144, 329-343.

Krieger et al. (2001) J. Virol. 4614-4624

Kwong AD. et al., 1998, Antiviral Res., 40, 1-18

15 Lanford *et al.* **1994**, Virology 202, 606

Lesburg et al. 1999, Nat. Struct. Biol. 6: 937-943

Lohman et al. 1999, Science 285: 110-113

Lohman et al. (2001) J. Virol. 1437-1449

Love, R.A. et al., 1996, Cell, 87, 331-342

20 Martell et al. 1999 J. Clin. Microbiol. 37: 327-332

Mizutani et al. 1996, J. Virol. 70, 7219 – 7223

Moser et al., 1998, J. Virol. 72, 5318

Reed et al., 1997, J. Virol. 71, 7187

Sambrook et al., 1989, Molecular Cloning - A Laboratory Manual, Cold Spring

Harbor Labs.

Shimizu et al. 1993, PNAS, USA, 90, 6037 - 6041

Yanagi et al., 1999, Proc. Natl. Acad. Sci. USA, 96, 2291-95

Yao et al., 1997, Nature Structural Biology, 4, 463

Yem et al., 1998, Protein Science, 7, 837

30 Yoo et al. 1995, J. Virol., 69, 32 – 38

CLAIMS

- 1. A HCV polynucleotide molecule comprising a 5'-non translated region (NTR) wherein guanine at position 1 is substituted for adenine.
- 2. A HCV self-replicating polynucleotide comprising:
 - a 5'-NTR consisting of ACCAGC (SEQ ID NO. 8);
 - a HCV polyprotein region coding for a HCV polyprotein; and
 - a 3'-NTR region.
- The HCV polynucleotide according to claim 2, wherein said polyprotein comprises one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T.
- 4. The HCV polynucleotide encoding a polyprotein comprising one or more of the amino acid substitution as defined in claim 3, and further comprising the amino acid substitution E(1202)G.
- 5. The HCV polynucleotide according to claim 3, wherein said substitution is a G2042C or a G2042R mutation.
- 6. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: K(1691)R; and G(2042)C.
- 7. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; and P(2166)L.
- 8. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; L(2155)P; and P(2166)L.
- 9. The HCV polynucleotide according to claim 3, wherein said substitution is selected

- from the group consisting of: E(1202)G; I(1984)V; G(2042)C; and M(2992)T.
- 10. The HCV polynucleotide according to claim 3, wherein said substitution is selected from the group consisting of: S(1148)G; E(1202)G; L(1701)F; G(2042)R; and S(2404)P.
- 11. The HCV polynucleotide according to claim 2, wherein said polynucleotide is a RNA molecule encoded by the DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- 12. The HCV polynucleotide according to claim 2, wherein said polynucleotide is a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- 13. An expression vector comprising a DNA form of the polynucleotide according to claim 2, operably linked to a promoter.
- 14. A host cell transfected with the self-replicating polynucleotide molecule according to claim 2.
- 15. A host cell according to claim 14, wherein the host cell is a eukaryotic cell line.
- 16. A host cell according to claim 15, wherein said eukaryotic cell line is a hepatic cell line.
- 17. A host cell according to claim 16, wherein said hepatic cell line is Huh-7.
- 18. A RNA replication assay comprising the steps of:
 - a) incubating the host cell according to claim 14 under conditions suitable for RNA replication;
 - b) isolating the total cellular RNA from the cells; and
 - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
- 19. The assay according to claim 18, wherein the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.

WO 02/052015 PCT/CA01/01843

- 20. The assay according to claim 18, wherein said polynucleotide encodes for a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.
- 21. A method for testing a compound for inhibiting HCV replication, including the steps of:
 - a) carrying step a) according to claim 18, in the presence or absence of the compound;
 - b) isolating the total cellular RNA from the cells; and
 - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
 - d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor,

wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

- 22. The method according to claim 21, wherein said cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.
- 23. A HCV polynucleotide molecule comprising:
 - a 5'-NTR region;
 - a HCV polyprotein region coding for a HCV polyprotein comprising one or more amino acid substitution selected from the group consisting of: R(1135)K; S(1148)G; S(1560)G; K(1691)R; L(1701)F; I(1984)V; T(1993)A; G(2042)C; G(2042)R; S(2404)P; L(2155)P; P(2166)L and M(2992)T; and a 3'-NTR region.
- 24. The HCV self-replicating polynucleotide encoding a polyprotein comprising the any one of the amino acid substitutions as defined in claim 24, further comprising the amino acid substitution E(1202)G.
- 25. The polynucleotide according to claim 24, wherein said substitution is a G2042C or a G2042R mutation.
- 26. The HCV polynucleotide according to claim 24, wherein said substitution is selected

- from the group consisting of: K(1691)R; and G(2042)C.
- 27. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; and P(2166)L.
- 28. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: R(1135)K; S(1560)G; K(1691)R; T(1993)A; G(2042)C; L(2155)P; and P(2166)L.
- 29. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: E(1202)G; I(1984)V; G(2042)C; and M(2992)T.
- 30. The HCV polynucleotide according to claim 24, wherein said substitution is selected from the group consisting of: S(1148)G; E(1202)G; L(1701)F; G(2042)R; and S(2404)P.
- 31. The HCV polynucleotide according to claim 24, wherein said molecule is a RNA molecule encoded by the DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- 32. The HCV polynucleotide according to claim 24, wherein said molecule is a DNA molecule selected from the group consisting of: SEQ ID NO. 2, 4, 5, 6, 7, 24 and 25.
- 33. An expression vector comprising a DNA form of the polynucleotide according to claim 24, operably linked to a promoter.
- 34. A host cell transfected with the self-replicating polynucleotide according to claim 24.
- 35. A host cell according to claim 34, wherein the host cell is a eukaryotic cell line.
- 36. A host cell according to claim 35, wherein said eukaryotic cell line is a hepatic cell line.
- 37. A host cell according to claim 36, wherein said hepatic cell line is Huh-7.

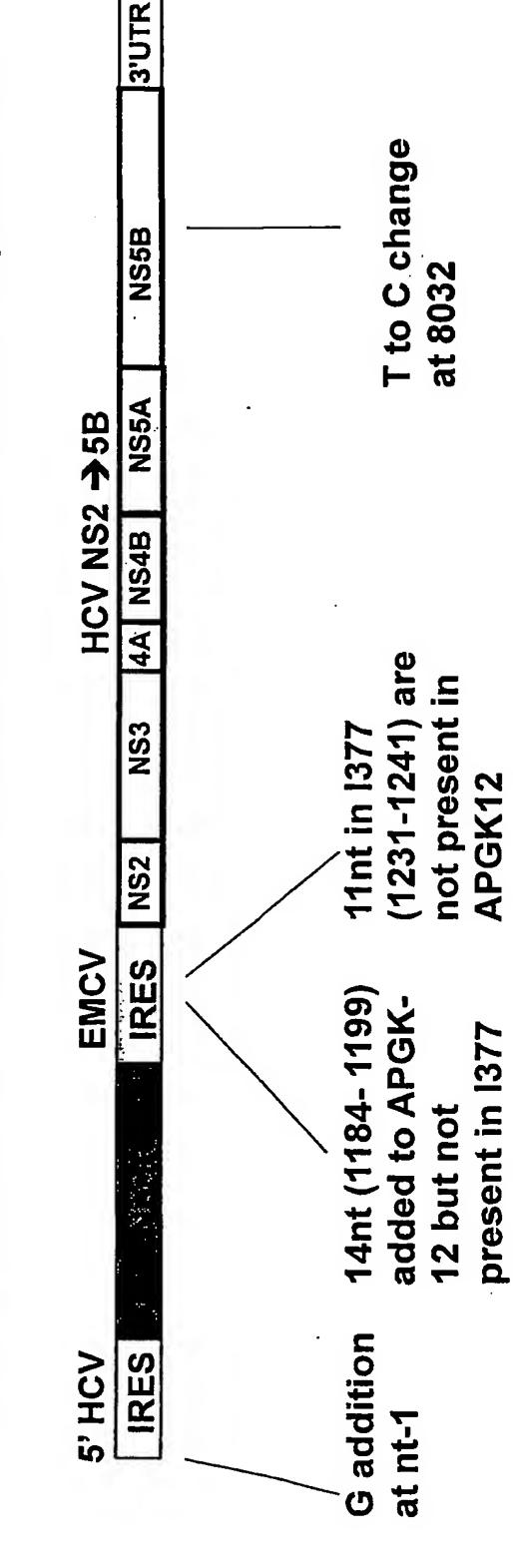
- 38. A RNA replication assay comprising the steps of:
 - incubating the host cell according to claim 34 under conditions suitable for RNA replication;
 - isolating the total cellular RNA from the cells; and analyzing the RNA so as to measure the amount of HCV RNA replicated.
- 39. The assay according to claim 38, wherein the analysis of RNA levels in step c) is carried out by amplifying the RNA by real-time RT-PCR analysis using HCV specific primers so as to measure the amount of HCV RNA replicated.
- 40. The assay according to claim 38, wherein said polynucleotide encodes for a reporter gene, and the analysis of RNA levels in step c) is carried out by assessing the level of reporter expressed.
- 41. A method for testing a compound for inhibiting HCV replication, including the steps of:
 - a) carrying step a) according to claim 38, in the presence or absence of the compound;
 - b) isolating the total cellular RNA from the cells; and
 - c) analyzing the RNA so as to measure the amount of HCV RNA replicated.
 - d) comparing the levels of HCV RNA in cells in the absence and presence of the inhibitor,

wherein reduced RNA levels is indicative of the ability of the compound to inhibit replication.

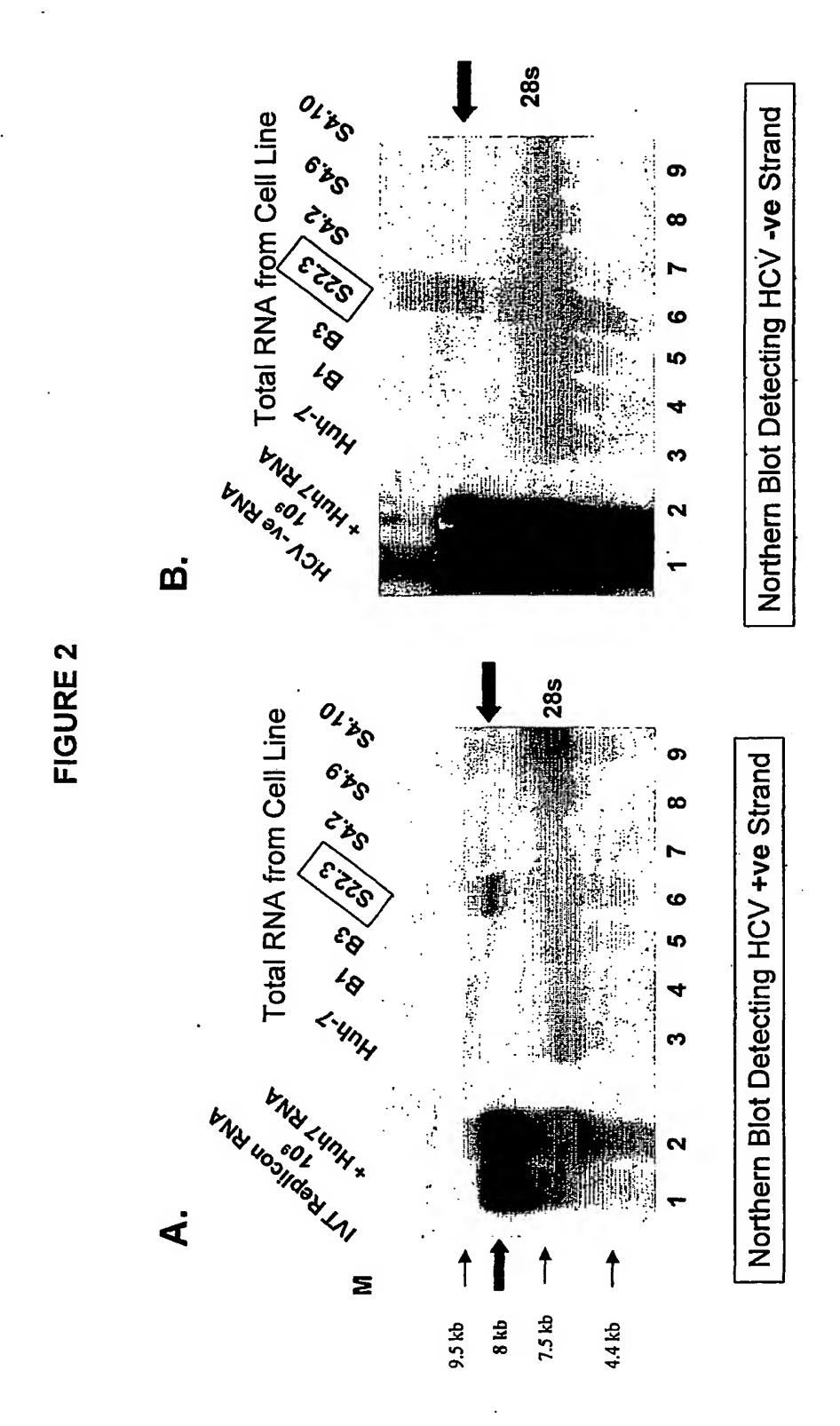
42. The method according to claim 41, wherein said cell line is incubated with the test compound for about 3-4 days at a temperature of about 37°C.

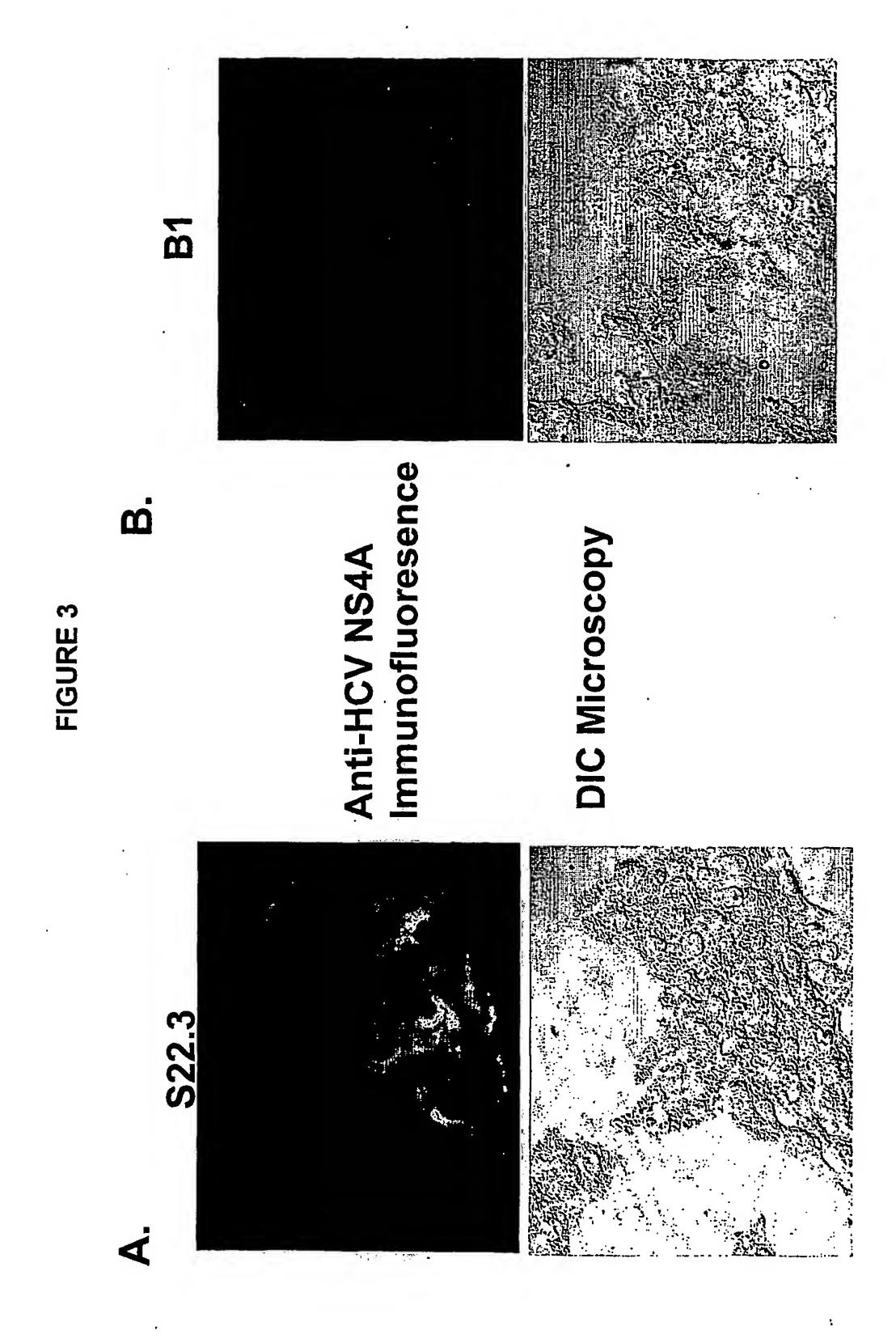
1/9

RNA compared to 1377/NS2-3' Replicon RNA APGK12 (SEQ ID NO 1) Replicon

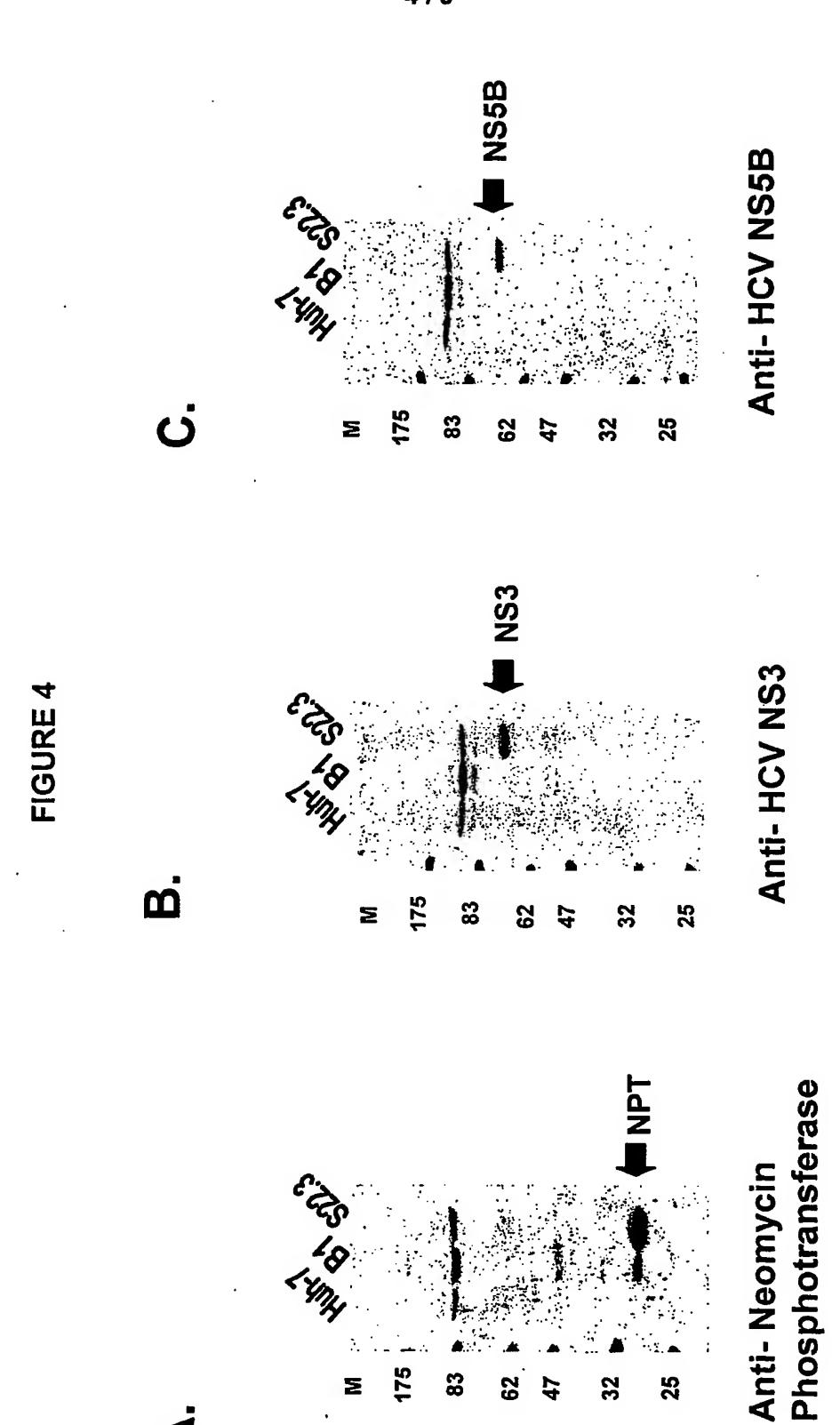


SUBSTITUTE SHEET (RULE 26)









SUBSTITUTE SHEET (RULE 26)

5/9

FIGURE 5A

	S 22-3 SEQ ID NO 2	R3 SEQ 1D NO. 4	R3-rep SEQ ID NO. 7	R7 SEQ ID NO. 5	R16 SEQ ID NO 6
5'end - FIRST nt (HCV IRES)	*G (nt 1) A	G (nt 1) A	•	•	G (nt 1) A
Neo	•	A (nt 481) G	•		· :
EMCV IRES		A (nt 1739) G		•	
NS.2		•	•	•	·
, NS 3	•	G (nt 2778) A A (nt 2840) C A (nt 4062) G	T (nt 2509) C G (nt 2778) A A (nt 2840) C T (nt 3574) C A (nt 4052) G	A (nt 2935) G A (nt 2979) G	A (nt 2816) G A (nt 2979) G
NS 4A	A (nt 4446) R	A (nt 4448) G	C (nt 4387) T A (nt 4446) G C (nt 4507) T	•	C (nt 4475) T
NS 4B		T (nt 4855) C	T (nt 4865) C		•
NS 5A	G (nt 6498) T A (nt 6268) R	A (nt 5351) G G (nt 5498) T G (nt 5659) A C (nt 5871) T A (nt 6268) G	A (nt 5351) G G (nt 5498) T G (nt 5659) A T (nt 5838) C C (nt 5871) T A (nt 5115) G	A (nt 5324) G G (nt 5498) T · T (nt 6001) C	G (nt 5498) C T (nt 6320) C T (nt 6584) C
NS 5B	•	A (nt 6662) G		C (nt 7252) T T (nt 8349) C	
3'end - last 98 nt	·		: ,		•

*first nt = G from HCV ires

FIGURE 5B

6/9

G (2042) R S (2404) P L (1701) F R16 SEQ S (1148) G E (1202) G **1D NO. 6** G (nt 1) A R7 SEQ ID G (2042) C E (1202) G M (2992) T I (1984) V NO. 5 ٠. R3 Rep SEQ ID R (1135) K S (1560) G G (2042) C K (1691) R L (2155) P T (1993) A P (2166) L R3 SEQ ID R (1135) K S (1560) G G (2042) C K (1691) R T (1993) A G (nt 1) A P (2166) L NO. 4 • ix K/R S 22-3 SEQ ID O 4 NO. 2 G (nt 1) G (2042) K (1691) mi 5'end - FIRST nt (HCV IRES) 3'end - last 98 nt NS 4A NS 4B NS 5A NS₂ NS 3 SZ

first a.a. of NS2 = 810

7/9

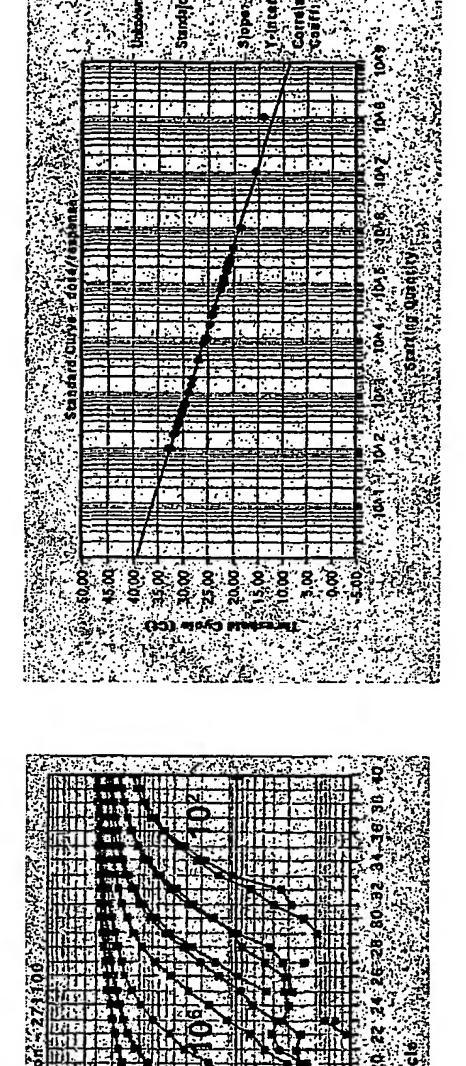
3'HCV UTR NS5B G(2042)C G(2042)C L(2155)P T(1993)A L(2155)P P(2166)L P(2166)L T(1993)A NS5A AMINO ACID SUBSTITUTIONS NS4B HCV NS2→5B K(1691)R K(1691)R R(1135)K R(1135)K S(1560)G S(1560)G FIGURE 6 NS3 NS2 **EMCV** IRES NeoR SEQ ID NO 24 SEQ ID NO 25 SEQ ID NO 1 SEQ ID NO 7 5' HCV IRES CLONE APGK-12 R3 rep G (nt1) A (nt1) G(nt1) A(nt1) 2000000cfu/μg 1100000cfu/µg 86 cfu/µg

SUBSTITUTE SHEET (RULE 26)

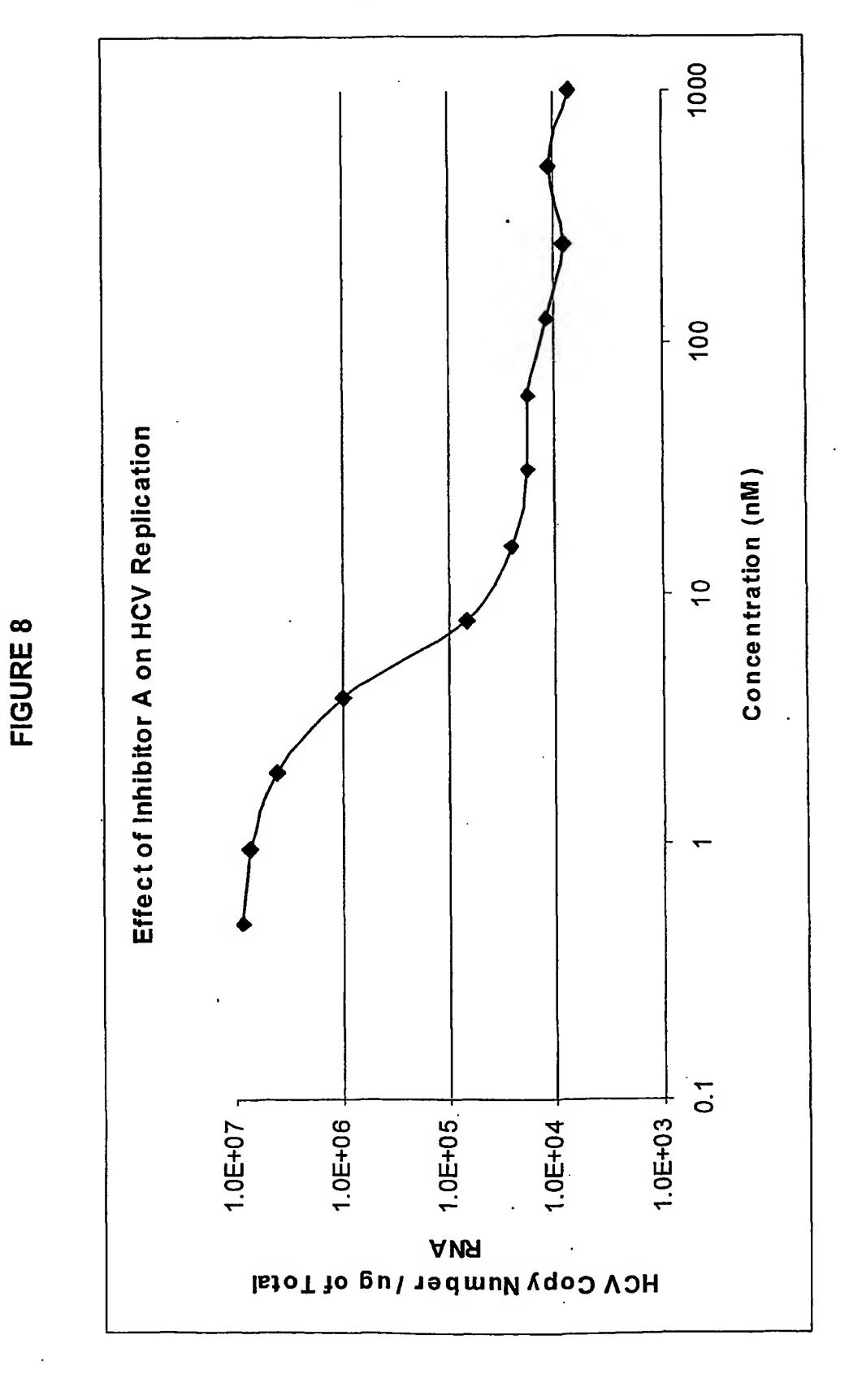
8/9

ICV-Replicon: RNA Quantification

FIGURE 7



Ct = Threshold cycle



SUBSTITUTE SHEET (RULE 26)

1/93

SEQUENCE LISTING

```
<110> BOEHRINGER INGELHEIM (CANADA) LTD.
<120> SELF REPLICATING RNA MOLECULE FROM
  HEPATITIS C VIRUS
<130> 13/083
<150> 60/257,857
<151> 2000-12-22
<160> 25
<170> FastSEQ for Windows Version 4.0
<210> 1
<211> 8639
<212> DNA
<213> HCV
<220>
<221> CDS
<222> (1803)...(8408)
<400> 1
ggccagcccc cgattggggg cgacactcca ccatagatca ctcccctgtg aggaactact 60
gtcttcacgc agaaagcgtc tagccatggc gttagtatga gtgtcgtgca gcctccagga 120
cccccctcc cgggagagcc atagtggtct gcggaaccgg tgagtacacc ggaattgcca 180
ggacgaccgg gtcctttctt ggatcaaccc gctcaatgcc tggagatttg ggcgtgcccc 240
cgcgagactg ctagccgagt agtgttgggt cgcgaaaggc cttgtggtac tgcctgatag 300
ggtgcttgcg agtgccccgg gaggtctcgt agaccgtgca ccatgagcac gaatcctaaa 360
cctcaaagaa aaaccaaagg gcgcgccatg attgaacaag atggattgca cgcaggttct 420
ccggccgctt gggtggagag gctattcggc tatgactggg cacaacagac aatcggctgc 480
tetgatgeeg cegtgtteeg getgteageg caggggegee eggttetttt tgteaagace 540
gacctgtccg gtgccctgaa tgaactgcag gacgaggcag cgcggctatc gtggctggcc 600
acgacgggcg ttccttgcgc agctgtgctc gacgttgtca ctgaagcggg aagggactgg 660
ctgctattgg gcgaagtgcc ggggcaggat ctcctgtcat ctcaccttgc tcctgccgag 720
aaagtateea teatggetga tgeaatgegg eggetgeata egettgatee ggetaeetge 780
ccattcgacc accaagcgaa acatcgcatc gagcgagcac gtactcggat ggaagccggt 840
cttgtcgatc aggatgatct ggacgaagag catcaggggc tcgcgccagc cgaactgttc 900
gccaggctca aggcgcgcat gcccgacggc gaggatctcg tcgtgaccca tggcgatgcc 960
tgcttgccga atatcatggt ggaaaatggc cgcttttctg gattcatcga ctgtggccgg 1020
ctgggtgtgg cggaccgcta tcaggacata gcgttggcta cccgtgatat tgctgaagag 1080
cttggcggcg aatgggctga ccgcttcctc gtgctttacg gtatcgccgc tcccgattcg 1140
cagcgcatcg ccttctatcg ccttcttgac gagttcttct gagttcgcgc ccagatgtta 1200
acagaccaca acggtttccc tctagcggga tcaattccgc ccccccct aacgttactg 1260
gccgaagccg cttggaataa ggccggtgtg cgtttgtcta tatgttattt tccaccatat 1320
tgccgtcttt tggcaatgtg agggcccgga aacctggccc tgtcttcttg acgagcattc 1380
ctaggggtct ttcccctctc gccaaaggaa tgcaaggtct gttgaatgtc gtgaaggaag 1440
cagttcctct ggaagcttct tgaagacaaa caacgtctgt agcgaccctt tgcaggcagc 1500
ggaacccccc acctggcgac aggtgcctct gcggccaaaa gccacgtgta taagatacac 1560
ctgcaaaggc ggcacaaccc cagtgccacg ttgtgagttg gatagttgtg gaaagagtca 1620
aatggctctc ctcaagcgta ttcaacaagg ggctgaagga tgcccagaag gtaccccatt 1680
gtatgggatc tgatctgggg cctcggtgca catgctttac atgtgtttag tcgaggttaa 1740
aaaacgtcta ggccccccga accacgggga cgtggttttc ctttgaaaaa cacgataata 1800
```

						gca g Ala <i>F</i>				_	_	_	_	1847
	_			-		ttg Leu		_			_		gct Ala	1895
						caa Gln					_		_	1943
_						ccc				 	_	_		1991
-				_	-	gcg Ala 70							_	2039
						ata Ile						_	_	2087
						tac Tyr	_	-	_	-			cgt Arg	2135
_	-	_	_	_		aag Lys							atg Met	2183
_		-	_	_	_	gca Ala			_	_			_	2231
						tgg Trp 150							gcg Ala	2279
		- 			_	gtc Val		_			_	_		2327
,						gcg Ala							ctg Leu	2375
						gjå aaa		_	_	_		_		2423
agc Ser						tgg Trp								2471

	caa Gln 225	_	_		_			_		_			_		aca Thr	2519
	cgg						=						_			2567
	aca Thr															2615
	tat Tyr												_			2663
_	acc Thr	_					•							=		2711
	ccc Pro 305											_	_	_	tcg Ser	2759
	ctt Leu												-	_		2807
	Gly												_			2855
	aag Lys															2903 ·
gtg 'Val	Gly															2951
	gac Asp 385	Phe		Pro	Val	Glu		Met	Glu	Thr						2999
	ttc Phe							_			_	_				3047
	gcc Ala															3095
	gct Ala															3143

		_		-	acc Thr									_	_		3191
	_				aac Asn										=		3239
					tac Tyr												3287
					gcc Ala 500							_	_	- -			3335
					act Thr												3383
•	Glu	Thr	Ala 530	Gly	Ala	Arg	Leu	Val 535	Val	Leu	Ala	Thr	Ala 540	Thr	Pro		3431
	Gly	Ser 545	Val	Thr	gtg Val	Pro	His 550	Pro	Asn	Ile	Glu	Glu 555	Val	Ala	Leu	Ser	3479
	Ser 560	Thr	Gly	Glu	atc Ile	Pro 565	Phe	Tyr	Gly	Lys	Ala 570	Ile	Pro	Ile	Glu	Thr 575	3527
	Ile	Lys	Gly	Gly	agg Arg 580	His	Leu	Ile	Phe	Cys 585	His	Ser	Lys	Гуs	Lys 590	Cys	3575
	Asp	Glu	Leu	Ala 595	gcg Ala	Lys	Leu	Ser	Gly 600	Leu	Gly	Leu	Asn	Ala 605	Val	Ala	3623
	Tyr	Tyr	Arg 610	Gly	ctt Leu	Asp	Val	Ser 615	Val	Ile	Pro	Thr	Ser 620	Gly	Asp	Val	3671
	Ile	Val 625	Val	Ala	acg Thr	Asp	Ala 630	Leu	Met	Thr	Gly	Phe 635	Thr	Gly	Asp	Phe	3719
	Asp 640	Ser	Val	Ile	gac Asp	Cys 645	Asn	Thr	Cys	Val	Thr 650	Gln	Thr	Val	Asp	Phe 655	3767
				Pro	acc Thr 660								_			_	3815

	_	_				_			_		_	ggt Gly				-	3863
												ccc Pro			_		3911
												ggc Gly 715		-			3959
C	_		_		-				_			cgg Arg	_				4007
							-	-	-		_	gag Glu				-	4055
7	al	Phe	Thr	Gly 755	Leu	Thr	His	Ile	Asp 760	Ala	His	ttc Phe	Peń	Ser 765	Gln	Thr	4103
Ι	yys	Gln	Ala 770	Gly	Asp	Asn	Phe	Pro 775	Tyr	Leu	Val	gca Ala	Tyr 780	Gln	Ala	Thr	4151
V	al	Cys 785	Ala	Arg	Ala	Gln	Ala 790	Pro	Pro	Pro	Ser	Trp 795	Asp	Gln	Met	_	4199
I 8	00 'Ys	Cys	Leu	Ile	Arg	Leu 805	Lys	Pro	Thr	Leu	His 810	ej aaa	Pro	Thr	Pro	Leu 815	4247
I	eu	Tyr	Arg	Leu	Gly 820	Ala	Val	Gln	Asn	Glu 825	Val	act Thr	Thr	Thr	His 830	Pro	4295
I	le	Thr	Lys	Tyr 835	Ile	Met	Ala	Cys	Met 840	Ser	Ala	gac Asp	Leu	Glu 845	Val	Val	4343
	_		_		_		Val					gca Ala					4391
												ggc Gly 875					4439
S	er 80	gga Gly	aag Lys	ccg Pro	gcc Ala	atc Ile 885	att Ile	ccc Pro	gac Asp	agg Arg	gaa Glu 890	gtc Val	ctt Leu	tac Tyr	cgg Arg	gag Glu 895	4487

	_		_	_	gag Glu	_			_					_	cag Gln	4535
	_	_		-	gaa Glu				_	_	_			_	ctg Leu	4583
		_			caa Gln			_		_				_		4631
_					gaa Glu	_				_		_				4679
	_	-			tat Tyr 965		_		_			_			aac Asn 975	4727
					ctg Leu		_									4775
				His	acc Thr		_		Asn		_			Trp	gtg Val	4823
_	_		Leu	-	cct Pro		_	Ala	_		_		Val		gcc Ala	4871
		Ala			gct. Ala	Val		_		_			_			4919
ata						1030)			CLY	1035	_	_			
	Asp	att			ggt Gly 1045	tat Tyr	gga	gca	aaa	gtg	1035 gca Ala	ggc		ctc	gtg · Val 1055	4967
Val 1040 gcc	Asp) ttt	att Ile aag	Leu gtc	Ala atg	Gly 1045 agc Ser	tat Tyr	gga Gly gag	gca Ala atg	ccc GJÅ aaa	gtg Val 1050 tcc ser	gca Ala	gag ggc ggc	Ala gac	ctc Leu ctg	Val 1055 gtt Val	4967 5015
Val 1040 gcc Ala aac	Asp ttt Phe cta	att Ile aag Lys	Leu gtc Val	Ala atg Met 1060 gct Ala	Gly 1045 agc Ser	tat Tyr ggc Gly	gga Gly gag Glu	gca Ala atg Met	ggg Gly ccc Pro 1069	gtg Val 1050 tcc ser	gca Ala acc Thr	gag Glu gtc	Ala gac Asp	ctc Leu ctg Leu 1070	Val 1055 gtt Val O	
Val 1040 gcc Ala aac Asn	Asp ttt Phe cta Leu	att Ile aag Lys ctc Leu	gtc Val cct Pro 1075	Ala atg Met 1060 gct Ala ata	agc Ser	tat Tyr ggc Gly ctc Leu	gga Gly gag Glu tcc ser	gca Ala atg Met cct Pro 1080	ggg Gly ccc pro 1069	gtg Val 1050 tcc ser gcc Ala	gca Ala acc Thr cta Leu	gag Glu gtc Val	gac Asp gtc Val 1089 gag Glu	ctc Leu ctg Leu 1070 ggg Gly	Val 1055 gtt Val gtc Val gct	5015

gtc tcc ccc acg Val Ser Pro Thr 1120					17
act cag atc ctc Thr Gln Ile Leu	Ser Ser Leu		Gln Leu Leu L	— — — —	5
cac cag tgg atc His Gln Trp Ile 1155	Asn Glu Asp		Pro Cys Ser G		13
cta aga gat gtt Leu Arg Asp Val 1170				at ttc aag 535 sp Phe Lys	51
acc tgg ctc cag Thr Trp Leu Gln 1185		Leu Pro Arg			9
ttc tca tgt caa Phe Ser Cys Gln 1200		_			£7
atg caa acc acc Met Gln Thr Thr			Ile Thr Gly H) 5
aac ggt tcc atg Asn Gly Ser Met 1235	Arg Ile Val		Thr Cys Ser A	-	13
cat gga aca ttc His Gly Thr Phe 1250)1
tcc ccg gcg cca Ser Pro Ala Pro 1265		Arg Ala Leu			19
gag tac gtg gag Glu Tyr Val Glu 1280	Val Thr Arg				37
atg acc act gac Met Thr Thr Asp			Gln Val Pro A		}5
ttc ttc aca gaa Phe Phe Thr Glu 1315	Val Asp Gly		His Arg Tyr A		33
tgc aaa ccc ctc Cys Lys Pro Leu 1330					31

	•		Ser G	_	_	gag ccc gaa Glu Pro Glu 1355		
_	Val Leu				Asp Pro	tcc cac att Ser His Ile 1370	Thr Ala G	_
	-		J Leu A			ccc ccc tcc Pro Pro Ser		
	_			er Ala		ttg aag gca Leu Lys Ala		
	_	Asp Ser		_	Asp Leu	atc gag gcc Ile Glu Ala 142	Asn Leu L	_
			Gly G			cgc gtg gag Arg Val Glu 1435		
	Val Val					ctc caa gcg Leu Gln Ala 1450	Glu Glu A	
			. Val P			ctg cgg agg Leu Arg Arg		
				le Trp	_	ccg gat tac Pro Asp Tyr		·
-		Ser Tr	_	_	Asp Tyr	gtc cct cca Val Pro Pro 150	Val Val H	
_			Pro A	-	_	ccg ata cca Pro Ile Pro 1515		 -
	Lys Arg	_				acc gtg tct Thr Val Ser 1530	Ser Ala I	
			Lys T			tcc gaa tcg Ser Glu Ser		
		_		la Ser		cag ccc tcc Gln Pro Ser		-
	•				•			

Asp Ala Gly Se 1570	gac gtt gag Asp Val Glu			Pro Leu Glu	551
ggg gag ccg ggg Gly Glu Pro Gly 1585		Leu Ser Asp			599
agc gag gag gc Ser Glu Glu Al 1600		-			647
tgg aca ggc gc Trp Thr Gly Ala			Ala Glu Glu		695
ccc atc aat gc Pro Ile Asn Ala 16	Leu Ser Asn	-	-		743
tat gct aca ac Tyr Ala Thr Th 1650				Lys Val Thr	791
ttt gac aga ct Phe Asp Arg Le 1665		Asp Asp His			839
gag atg aag gc	aag gcg tcc Lys Ala Ser				887
1680	1685	TIEL VOIL ENGL	1690	1695	
	1685 aag ctg acg	ccc cca cat	tcg gcc aga Ser Ala Arg	tct aaa ttt 695	935
gag gaa gcc tg Glu Glu Ala Cy	aag ctg acg Lys Leu Thr 1700 aag gac gtc Lys Asp Val	ccc cca cat Pro Pro His 170 cgg aac cta	tcg gcc aga Ser Ala Arg tcc agc aag	tct aaa ttt 695 Ser Lys Phe 1710 gcc gtt aac 69	
gag gaa gcc tg Glu Glu Ala Cy ggc tat ggg gc Gly Tyr Gly Ala 17 cac atc cgc tc His Ile Arg Se	aag ctg acg Lys Leu Thr 1700 aag gac gtc Lys Asp Val	ccc cca cat Pro Pro His 170 cgg aac cta Arg Asn Leu 1720 gac ttg ctg Asp Leu Leu	tcg gcc aga Ser Ala Arg tcc agc aag Ser Ser Lys gaa gac act Glu Asp Thr	tct aaa ttt 695 Ser Lys Phe 1710 gcc gtt aac 69 Ala Val Asn 1725 gag aca cca 76 Glu Thr Pro	
gag gaa gcc tg Glu Glu Ala Cy ggc tat ggg gc Gly Tyr Gly Ala 17 cac atc cgc tc His Ile Arg Se	aag ctg acg Lys Leu Thr 1700 aag gac gtc Lys Asp Val 5 gtg tgg aag Val Trp Lys atc atg gca	ccc cca cat Pro Pro His 170 cgg aac cta Arg Asn Leu 1720 gac ttg ctg Asp Leu Leu 1735 aaa aat gag Lys Asn Glu	tcg gcc aga Ser Ala Arg tcc agc aag Ser Ser Lys gaa gac act Glu Asp Thr 1746	tct aaa ttt 695 Ser Lys Phe 1710 gcc gtt aac 69 Ala Val Asn 1725 gag aca cca 70 Glu Thr Pro	983
gag gaa gcc tg Glu Glu Ala Cy ggc tat ggg gcc Gly Tyr Gly Ala 17: cac atc cgc tcc His Ile Arg Sec 1730 att gac acc acc Ile Asp Thr Th	aag ctg acg Lys Leu Thr 1700 aag gac gtc Lys Asp Val gtg tgg aag Val Trp Lys atc atg gca Tle Met Ala 175 cgc aag cca	ccc cca cat Pro Pro His 170 cgg aac cta Arg Asn Leu 1720 gac ttg ctg Asp Leu Leu 1735 aaa aat gag Lys Asn Glu 0 gct cgc ctt	tcg gcc aga Ser Ala Arg tcc agc aag Ser Ser Lys gaa gac act Glu Asp Thr 1746 gtt ttc tgc Val Phe Cys 1755 atc gta ttc	tct aaa ttt 695 Ser Lys Phe 1710 gcc gtt aac 69 Ala Val Asn 1725 gag aca cca 70 Glu Thr Pro gtc caa cca 70 Val Gln Pro cca gat ttg 73	983 031

.10/93

	-	gga ttc caa tac tct cct 72 Gly Phe Gln Tyr Ser Pro 1805	223
		tgg aaa gcg aag aaa tgc 72 Trp Lys Ala Lys Lys Cys 1820	271
		ttt gac tca acg gtc act 73 Phe Asp Ser Thr Val Thr 1835	319
_		tac caa tgt tgt gac ttg 73 Tyr Gln Cys Cys Asp Leu 1850 1855	367
	Gln Ala Ile Arg Ser	Leu Thr Glu Arg Leu Tyr	115
		cag aac tgc ggc tat cgc 74 Gln Asn Cys Gly Tyr Arg 1885	463
	_	agc tgc ggt aat acc ctc 75 Ser Cys Gly Asn Thr Leu 1900	511
		cga gct gcg aag ctc cag 75 Arg Ala Ala Lys Leu Gln 1915	559
		ctt gtc gtt atc tgt gaa 76 Leu Val Val Ile Cys Glu 1930 1935	507
	Glu Asp Glu Ala Ser	Leu Arg Ala Phe Thr Glu	6,55
		gac ccg ccc aaa cca gaa 77 Asp Pro Pro Lys Pro Glu 1965	703
	_	tcc aat gtg tca gtc gcg 77 Ser Asn Val Ser Val Ala 1980	751
	- -	ctc acc cgt gac ccc acc 77 Leu Thr Arg Asp Pro Thr 1995	799
— — — — — — — — — — — — — — — — — — —		gct aga cac act cca gtc 78 Ala Arg His Thr Pro Val 2010 2015	847

11/93

	cta ggc aac at Leu Gly Asn Il 2020	_				Ala
Arg Met Ile	ctg atg act ca Leu Met Thr Hi 2035		Ser Ile			-
	aaa gcc cta ga Lys Ala Leu As				Cys Tyr	
_ _	ctt gac cta co Leu Asp Leu Pi 20			_		
•	tca ctc cat ac Ser Leu His Se 2085		- -	Glu Ile		-
	ctc agg aaa ct Leu Arg Lys Le 2100					Arg
His Arg Ala	aga agt gtc co Arg Ser Val Ar 2115		Leu Leu			
	tgt ggc aag ta Cys Gly Lys Ty				Arg Thr	
	act cca atc co Thr Pro Ile Pr 21					
_	gct ggt tac ag Ala Gly Tyr Se 2165			Tyr His		
	ccc cgc tgg tt Pro Arg Trp Pl 2180					Val
Gly Val Gly	atc tat cta ct Ile Tyr Leu Le 2195		Arg *	acggggag	jct aaaca	actcca 8428
	catcctgtt tttt	•				
	tttctcctt tttt					
	ctagtcacg gcta	_	aggtccgt	gagccgct	tg actgo	_
210> 2	ggcctctct gcag	jaluady t				8639

<210> 2

<211> 8642

<212> DNA

<213> HCV

<220>

```
<221> CDS
<222> (1802)...(8407)
<221> variation
<222> 6268
<223> r = a or g
<221> variation
<222> 4446
<223> r = a or g
<400> 2
accageceee gattggggge gacaeteeae catagateae teeeetgtga ggaactaetg 60
tetteaegea gaaagegtet ageeatggeg ttagtatgag tgtegtgeag ceteeaggae 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg teettettg gateaacccg eteaatgeet ggagatttgg gegtgeecc 240
gegagaetge tageegagta gtgttgggte gegaaaggee ttgtggtaet geetgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtatecat catggetgat geaatgegge ggetgeatae gettgateeg getaeetgee 780
cattegacca ccaagegaaa categeateg agegageaeg taeteggatg gaageeggte 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggeteaa ggegegeatg ceegaeggeg aggatetegt egtgaeeeat ggegatgeet 960
gcttgccgaa tatcatggtg gaaaatggcc gcttttctgg attcatcgac tgtggccggc 1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc 1140
agegeatege ettetatege ettettgaeg agttettetg agttegegee eagatgttaa 1200
cagaccacaa cggtttccct ctagcgggat caattccgcc cccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380
taggggtctt tcccctctcg ccaaaggaat gcaaggtctg ttgaatgtcg tgaaggaagc 1440
agttcctctg gaagcttctt gaagacaaac aacgtctgta gcgacccttt gcaggcagcg 1500
gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggctctcc tcaagcgtat tcaacaaggg gctgaaggat gcccagaagg taccccattg 1680
tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740
aaacgtctag gcccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
  Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly
   1
                   5
                                       10
                                                           15
ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg
                                                                   1897
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg
             20
                                 25
                                                     30
ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg
                                                                   1945
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu
         35
                             40
                                                 45
caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc
                                                                  1993
Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val
     50
                         55
                                             60
```

_		ctc Leu		-	_											2041
		ttg Leu												_		2089
_	_	aaa Lys	_						_						gca Ala	2137
		ctg Leu 115			-	-						_		_	_	2185
		aag Lys														2233
		ctg Leu														2281
															acc Thr	2329
		gca Ala														2377
		gcc Ala 195														2425 ·
		GJA aaa	_									_				2473
		acg Thr			Leu	Leu		Cys	Ile			_				2521
		agg Arg														2569
aca Thr	caa Gln	tct Ser	ttc Phe 260	ctg Leu	gcg Ala	acc Thr	tgc Cys	gtc Val 265	aat Asn	ggc	gtg Val	tgt Cys	tgg Trp 270	act Thr	gtc Val	2617
		ggt Gly 275														2665

_	_					_		_			gtc Val 300				gcg Ala	2713
			_	_					=		tgc Cys				_	2761
				_		_	_		_	_	ecg Pro	_ "				2809
-		-									ccc Pro				ttg Leu	2857
	_			_	_	-					tcg Ser			_	gtg Val	2905
ggc		_		_	_	_		_		_	gtt Val 380					2953
		_									atg Met			_	gtc Val 400	3001
									-		cag Gln			_	gtg Val	3049
											agc Ser				ccg Pro	3097
-			_	-					-		gtc Val	-		_	tcc Ser	3145
											tct Ser 460				ggt Gly	3193
											atc Ile		-		_	3241
											gcc Ala	_	_		_	3289
											gag Glu	_				3337

Asp	tcg Ser				_							_				3385
	gct Ala 530								_		_			_	gga Gly	3433
-	gtc Val											_	_		agc Ser 560	3481
	gga Gly															3529
	Gly aaa		-								_	_		_	gat Asp	3577
	ctc Leu			_	_							_	· -	_		3625
	cgg Arg 610			_			_				_		_	_	att Ile	3673
	gta		_							_			-		gac	3721
625	Val	Ala	Thr	Asp	630	ren	Met		GIY	635	1111	Q ₁	Asp	PHE	Asp 640	
625 tca		atc	gac	tgc	630 aat	aca	tgt	gtc	acc	635 cag	aca	gtc	gac	ttc	640 agc	3769
625 tca Ser	gtg Val	atc Ile ccg	gac Asp	tgc Cys 645	630 aat Asn	aca Thr	tgt Cys gag	gtc Val	acc Thr 650	635 cag Gln	aca Thr	gtc Val cca	gac Asp	ttc Phe 655 gac	agc Ser	3769 3817
tca Ser ctg Leu	gtg Val gac Asp	atc Ile ccg Pro	gac Asp acc Thr 660	tgc Cys 645 ttc Phe	aat Asn acc Thr	aca Thr att Ile	tgt Cys gag Glu	gtc Val acg Thr 665	acc Thr 650 acg Thr	cag Gln acc Thr	aca Thr gtg Val	gtc Val cca Pro	gac Asp caa Gln 670	ttc Phe 655 gac Asp	agc Ser gcg Ala	
tca Ser ctg Leu gtg Val	gtg Val gac Asp	atc Ile ccg Pro cgc Arg 675	gac Asp acc Thr 660 tcg Ser	tgc Cys 645 ttc Phe cag Gln	aat Asn acc Thr cgg Arg	aca Thr att Ile cga Arg	tgt Cys gag Glu ggc Gly 680	gtc Val acg Thr 665 agg Arg	acc Thr 650 acg Thr act Thr	cag Gln acc Thr ggt Gly	aca Thr gtg Val agg Arg	gtc Val cca Pro ggc Gly 685	gac Asp caa Gln 670 agg Arg	ttc Phe 655 gac Asp atg Met	agc Ser gcg Ala ggc Gly	3817
tca ser ctg Leu gtg Val att Ile	gtg Val gac Asp tca ser tac	atc Ile ccg Pro cgc Arg 675 agg Arg	gac Asp acc Thr 660 tcg Ser ttt Phe	tgc Cys 645 ttc Phe cag Gln gtg Val	aat Asn acc Thr cgg Arg act Thr	aca Thr att Ile cga Arg cca Pro 695	tgt Cys gag Glu ggc Gly 680 gga Gly	gtc Val acg Thr 665 agg Arg gaa Glu	acc Thr 650 acg Thr cgg Arg	cag Gln acc Thr ggt Gly ccc Pro	aca Thr gtg Val agg Arg tcg ser 700 tgt	gtc Val cca Pro ggc Gly 685 ggc Gly	gac Asp caa Gln 670 agg Arg atg Met	ttc Phe 655 gac Asp atg Met ttc Phe	agc ser gcg Ala ggc Gly gat Asp	3817 3865

		_		gtc Val	_	_	_		_	_				_	_	4057	
	Thr			acc Thr			_	_					_		_	4105	
_	_		_	aac Asn				_	_	_		_	_	_		4153	
_	_		_	cag Gln	_				_				_		_	4201	
Cys	Leu	Ile	Arg	Cta Leu 805	Lys	Pro	Thr	Leu	His 810	Gly	Pro	Thr	Pro	Leu 815	Leu	4249	
Tyr	Arg	Leu	Gly 820	gcc Ala	Val	Gln	Asn	Glu 825	Val	Thr	Thr	Thr	His 830	Pro	Ile	4297	
Thr	Lys	Tyr 835	Ile	atg Met ctg	Ala	Cys	Met 840	Ser	Ala	Asp	Leu	Glu 845	Val	Val	Thr	4345 · 4393	
Ser	Thr 850	Trp	Val	Leu	Val	Gly 855	Gly	Val	Leu	Ala	Ala 860	Leu	Ala	Ala	Tyr	4441	
Cys 865	Leu	Thr	Thr	Gly	Ser 870	Val	Val	Ile	Val	Gly 875	Arg	Ile	Ile	Leu	Ser 880	4489	
Gly	Xaa	Pro	Ala	Ile 885	Ile	Pro	Asp	Arg	Glu 890	Val	Leu	Tyr	·Arg	Glu 895	Phe	4537	
Asp	Glu	Met	Glu 900	Glu	Cys	Ala	Ser	His 905	Leu	Pro	Tyr	Ile	Glu 910	Gln		4585	
Met	Gln	Leu 915	Ala	Ģlu caa	Gln	Phe	Lys 920	Gln	Lys	Ala	Ile	Gly 925	Leu	Leu	Gln	4633	
Thr	Ala 930	Thr	Lys	Gln gaa	Ala	Glu [*] 935	Ala	Ala	Ala	Pro	Val 940	Val	Glu	Ser	Lys	4681	
				Glu	_			_									

agc ggg ata Ser Gly Ile						729
gcg ata gca Ala Ile Ala			Ala Ser			777
acc acc caa Thr Thr Gln 99!	His Thr Leu					325
gcc caa ctt Ala Gln Leu 1010			Ser Ala I			373.
atc gct gga Ile Ala Gly 1025		Gly Ser Ile				921
gat att ttg Asp Ile Leu	gca ggt tat Ala Gly Tyr 1045	gga gca ggg Gly Ala Gly	gtg gca g Val Ala (1050	ggc gcg ctc Gly Ala Leu	gtg gcc 49 Val Ala 1055	969
ttt aag gtc Phe Lys Val			Ser Thr G		Val Asn	17
cta ctc cct Leu Leu Pro 1075	Ala Ile Leu	tcc cct ggc Ser Pro Gly 1080	gcc cta c Ala Leu V	gtc gtc ggg Val Val Gly 1085	gtc gtg 50 Val Val	065
tgc gca gcg Cys Ala Ala 1090			Gly Pro G			113
cag tgg atg Gln Trp Met 1105		Ile Ala Phe				L61
tcc ccc acg Ser Pro Thr	His Tyr Val	cct gag agc Pro Glu Ser	Asp Ala A	Ala Ala Arg	Val Thr	209
cag atc ctc Gln Ile Leu	tct agt ctt Ser Ser Leu 1140	acc atc act Thr Ile Thr 114	Gln Leu I	ctg aag agg Leu Lys Arg 1150	Leu His	257
cag tgg atc Gln Trp Ile 1155	Asn Glu Asp	tgc tcc acg Cys Ser Thr 1160	cca tgc t Pro Cys S	tcc ggc tcg Ser Gly Ser 1165	tgg cta 53 Trp Leu	305
aga gat gtt Arg Asp Val 1170	tgg gat tgg Trp Asp Trp	ata tgc acg Ile Cys Thr 1175	Val Leu 7	act gat ttc Thr Asp Phe 1180	aag acc 53 Lys Thr	353

tgg ctc cag tcc aag Trp Leu Gln Ser Lys 1185			
tca tgt caa cgt ggg Ser Cys Gln Arg Gly 120	Tyr Lys Gly Val		
caa acc acc tgc cca Gln Thr Thr Cys Pro 1220		Ile Thr Gly His	
tgt tcc atg agg atc Cys Ser Met Arg Ile 1235		_	Thr Trp His
gga aca ttc ccc att Gly Thr Phe Pro Ile 1250			
ccg gcg cca aat tat Pro Ala Pro Asn Tyr 1265			
tac gtg gag gtt acg Tyr Val Glu Val Thr 1289	Arg Val Gly Asp		
acc act gac aac gta Thr Thr Asp Asn Val 1300		Gln Val Pro Ala	_
ttc aca gaa gtg gat Phe Thr Glu Val Asp 1315			Pro Ala Cys
aaa ccc ctc cta cgg Lys Pro Leu Leu Arg 1330			
tac ctg gtt ggg tca Tyr Leu Val Gly Ser 1345	Gln Leu Pro Cys	Glu Pro Glu Pro	Asp Val Ala
gtg ctc act tcc atg Val Leu Thr Ser Met 136	Leu Thr Asp Pro	-	
gct aag cgt agg ctg Ala Lys Arg Arg Leu 1380		Pro Pro Ser Leu	•
tca gct agc cag ctg Ser Ala Ser Gln Leu 1395	- -		Cys Thr Thr

cgt cat gac tcc Arg His Asp Ser 1410		Asp Leu Ile			6073
cgg cag gag ato Arg Gln Glu Met 1425				-	6121
gta gta att ttg Val Val Ile Lev			ı Gln Ala Glu		6169
agg gaa gta tco Arg Glu Val Ser 146	Val Pro Ala	gag atc ctg Glu Ile Leu 1465	g cgg agg tcc Arg Arg Ser	agg aaa ttc Arg Lys Phe 1470	6217 [·]
cct cga gcg atg Pro Arg Ala Met 1475	ccc ata tgg Pro Ile Trp	gca cgc ccg Ala Arg Pro 1480	g gat tac aac Asp Tyr Asn 148	Pro Pro Leu	6265
ttr gag tcc tgg Xaa Glu Ser Trp 1490		Asp Tyr Val		- ·-	6313
tgt cca ttg ccg Cys Pro Leu Pro 1505					6361
aag agg acg gtt Lys Arg Thr Val			. Val Ser Ser		6409
gag ctc gcc aca Glu Leu Ala Thr 154	Lys Thr Phe				6457
agc ggc acg gca Ser Gly Thr Ala 1555				Asp Gly Asp	6505
gcg gga tcc gad Ala Gly Ser Asr 1570	Val Glu Ser		Met Pro Pro		6553
gag ccg ggg gat Glu Pro Gly Asp 1585	Pro Asp Leu				6601
gag gag gct agt Glu Glu Ala Ser	Glu Asp Val 1605	Val Cys Cys	s Ser Met Ser LO	Tyr Thr Trp 1615	6649
Thr Gly Ala Leu . 162	Ile Thr Pro				6697
atc aat gca ctg Ile Asn Ala Leu 1635				Leu Val Tyr	6745

		Thr		cgc Arg		_	Ser					Lys	_			6793
	Arg			gtc Val		Asp					Asp				gag Glu 1680	6841
·-	-		_	gcg Ala 1685	Ser				_	Lys						6889
				Leu			Pro		Ser					Phe	ggc Gly	6937
	_	_	Lys	gac Asp	_			Leu			_		Val		cac His	6985
		Ser		tgg Trp			Leu					Glu				7033
_	Thr			atg Met	_	Lys			-		Cys	_			gag Glu 1760	7081
				aag Lys 1765	Pro	_	_			Val			_	_		7129
				Glu					Tyr					Thr	ctc Leu	7177
			Val	atg Met				Tyr					Ser		gga Gly	7225
		Val		ttc Phe			Asn					Lys			ect Pro	7273
_	Gly	_	_	tat Tyr	_	Thr	_	_		-	Ser	_	_		gag Glu 1840	7321
				gtt Val 1845	Glu					Gln		_	_	_	Ala	7369
	_			cag Gln			Arg	_	Leu		-			Tyr		7417

GJA GJ aaa aa	ly :		Leu					Gly					Tyr	_		7465
tgc co Cys Ai		Ala					Thr					Asn				7513
tgt ta Cys Ty 1905						Ala					Ala					7561
tgc ac					Cys					Val						7609
gcg gg Ala Gl		Thr		Glu				_	Leu		_			Glu		7657
atg ad Met Th	hr 1		Tyr					Gly					Pro			7705
gac tt Asp Le		Glu					Cys			Asn		Ser	_	_		7753
gat go Asp Al 1985						Val					Arg					7801
ccc ct Pro Le		_			Ala					Arg					Asn	7849
tcc to Ser Tr		Leu		Asn					Ala					Ala		7897
atg at Met Il	le 1		Met					Ser					Gln			79 4 5
ctt ga Leu Gl 20		Lys					Gln				_	Cys				7993
gag co Glu Pr 2065				Leu		Gln					Leu					8041
gca tt Ala Ph			Leu		Ser			Pro		Glu					Ala	8089

tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His 2100 2105 2110	8137
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala 2115 2120 2125	8185
gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu 2130 2135 2140	8233
aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp 2145 2150 2155 2160	8281
ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg 2165 2170 2175	8329
gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Leu Ser Val Gly 2180 2185 2190	8377
gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca Val Gly Ile Tyr Leu Leu Pro Asn Arg * 2195 2200	8427
ggccaatagg ccatcctgtt tttttcccct tttttttt tttttttt tttttttt	8547
<210> 3 <211> 2201 <212> PRT <213> HCV	
<220> <221> VARIANT <222> 882 <223> Xaa is Lys or Arg	
<221> VARIANT <222> 1489 <223> Xaa is Leu	
<pre><400> 3 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly 1</pre>	
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg	
25 30 ,	
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu 35 40 45	
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu	

23 / 93

Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala Cys Met Leu Val Arg Lys Val Ala Gly Gly His Tyr Val Gln Met Ala Leu Met Lys Leu Ala Ala Leu Thr Gly Thr Tyr Val Tyr Asp His Leu Thr Pro Leu Arg Asp Trp Ala His Ala Gly Leu Arg Asp Leu Ala Val · Ala Val Glu Pro Val Val Phe Ser Asp Met Glu Thr Lys Val Ile Thr Trp Gly Ala Asp Thr Ala Ala Cys Gly Asp Ile Ile Leu Gly Leu Pro Val Ser Ala Arg Arg Gly Arg Glu Ile His Leu Gly Pro Ala Asp Ser Leu Glu Gly Gln Gly Trp Arg Leu Leu Ala Pro Ile Thr Ala Tyr Ser Gln Gln Thr Arg Gly Leu Leu Gly Cys Ile Ile Thr Ser Leu Thr Gly Arg Asp Arg Asn Gln Val Glu Gly Glu Val Gln Val Val Ser Thr Ala Thr Gln Ser Phe Leu Ala Thr Cys Val Asn Gly Val Cys Trp Thr Val Tyr His Gly Ala Gly Ser Lys Thr Leu Ala Gly Pro Lys Gly Pro Ile Thr Gln Met Tyr Thr Asn Val Asp Gln Asp Leu Val Gly Trp Gln Ala Pro Pro Gly Ala Arg Ser Leu Thr Pro Cys Thr Cys Gly Ser Ser Asp Leu Tyr Leu Val Thr Arg His Ala Asp Val Ile Pro Val Arg Arg Arg Gly Asp Ser Arg Gly Ser Leu Leu Ser Pro Arg Pro Val Ser Tyr Leu Lys Gly Ser Ser Gly Gly Pro Leu Leu Cys Pro Ser Gly His Ala Val Gly Ile Phe Arg Ala Ala Val Cys Thr Arg Gly Val Ala Lys Ala Val Asp Phe Val Pro Val Glu Ser Met Glu Thr Thr Met Arg Ser Pro Val Phe Thr Asp Asn Ser Ser Pro Pro Ala Val Pro Gln Thr Phe Gln Val Ala His Leu His Ala Pro Thr Gly Ser Gly Lys Ser Thr Lys Val Pro Ala Ala Tyr Ala Ala Gln Gly Tyr Lys Val Leu Val Leu Asn Pro Ser Val Ala Ala Thr Leu Gly Phe Gly Ala Tyr Met Ser Lys Ala His Gly Ile Asp Pro Asn Ile Arg Thr Gly Val Arg Thr Ile Thr Thr Gly Ala Pro Ile Thr Tyr Ser Thr Tyr Gly Lys Phe Leu Ala Asp Gly Gly Cys Ser Gly Gly Ala Tyr Asp Ile Ile Ile Cys Asp Glu Cys His Ser Thr Asp Ser Thr Thr Ile Leu Gly Ile Gly Thr Val Leu Asp Gln Ala Glu Thr Ala Gly Ala Arg Leu Val Val Leu Ala Thr Ala Thr Pro Pro Gly

24/93

Ser Val Thr Val Pro His Pro Asn Ile Glu Glu Val Ala Leu Ser Ser Thr Gly Glu Ile Pro Phe Tyr Gly Lys Ala Ile Pro Ile Glu Thr Ile Lys Gly Gly Arg His Leu Ile Phe Cys His Ser Lys Lys Lys Cys Asp Glu Leu Ala Ala Lys Leu Ser Gly Leu Gly Leu Asn Ala Val Ala Tyr Tyr Arg Gly Leu Asp Val Ser Val Ile Pro Thr Ser Gly Asp Val Ile Val Val Ala Thr Asp Ala Leu Met Thr Gly Phe Thr Gly Asp Phe Asp Ser Val Ile Asp Cys Asn Thr Cys Val Thr Gln Thr Val Asp Phe Ser Leu Asp Pro Thr Phe Thr Ile Glu Thr Thr Thr Val Pro Gln Asp Ala Val Ser Arg Ser Gln Arg Arg Gly Arg Thr Gly Arg Gly Arg Met Gly Ile Tyr Arg Phe Val Thr Pro Gly Glu Arg Pro Ser Gly Met Phe Asp Ser Ser Val Leu Cys Glu Cys Tyr Asp Ala Gly Cys Ala Trp Tyr Glu Leu Thr Pro Ala Glu Thr Ser Val Arg Leu Arg Ala Tyr Leu Asn Thr Pro Gly Leu Pro Val Cys Gln Asp His Leu Glu Phe Trp Glu Ser Val Phe Thr Gly Leu Thr His Ile Asp Ala His Phe Leu Ser Gln Thr Lys Gln Ala Gly Asp Asn Phe Pro Tyr Leu Val Ala Tyr Gln Ala Thr Val Cys Ala Arg Ala Gln Ala Pro Pro Pro Ser Trp Asp Gln Met Trp Lys Cys Leu Ile Arg Leu Lys Pro Thr Leu His Gly Pro Thr Pro Leu Leu Tyr Arg Leu Gly Ala Val Gln Asn Glu Val Thr Thr His Pro Ile Thr Lys Tyr Ile Met Ala Cys Met Ser Ala Asp Leu Glu Val Val Thr Ser Thr Trp Val Leu Val Gly Gly Val Leu Ala Ala Leu Ala Ala Tyr Cys Leu Thr Thr Gly Ser Val Val Ile Val Gly Arg Ile Ile Leu Ser Gly Xaa Pro Ala Ile Ile Pro Asp Arg Glu Val Leu Tyr Arg Glu Phe Asp Glu Met Glu Glu Cys Ala Ser His Leu Pro Tyr Ile Glu Gln Gly Met Gln Leu Ala Glu Gln Phe Lys Gln Lys Ala Ile Gly Leu Leu Gln Thr Ala Thr Lys Gln Ala Glu Ala Ala Ala Pro Val Val Glu Ser Lys Trp Arg Thr Leu Glu Ala Phe Trp Ala Lys His Met Trp Asn Phe Ile 960. Ser Gly Ile Gln Tyr Leu Ala Gly Leu Ser Thr Leu Pro Gly Asn Pro Ala Ile Ala Ser Leu Met Ala Phe Thr Ala Ser Ile Thr Ser. Pro Leu Thr Thr Gln His Thr Leu Leu Phe Asn Ile Leu Gly Gly Trp Val Ala

ALA					Pro		Ala						Gly	Ala	Gly
Ile			Ala		Val		5 Ser				1020 Gly		Val	Leu	Val
102	5				1030)				1035	· .	-			1040
Asp	Ile	Leu	Ala	_		_	Ala	_		Ala)	_			Val 1055	
Phe	Lys	Val	Met												Asn
	_											_	1070		
Leu	Leu		Ala 5				Pro 1080	_		Leu		Val 1085	_	Val	Val
Суз			Ile				His 5				Gly		Gly	Ala	Val
						Ile	Ala	Phe	Ala	Ser	_	_			Val
	5_														1120
				1125	5		Glu		1130)				1135	5
Gln	Ile	Leu				Thr				Leu		_	Arg 1150		His
Gln	Trp	Ile 1159		Glu	Asp	Cys			Pro	·	Ser	Gly 1165		Trp	Leu
Arg	Asp 1170	_	Trp			Ile 1175	Cys	Thr		Leu	Thr 1180	_	Phe	Lys	Thr
Trp							Pro						Pro	Phe	Phe
118				_							5				1200
Ser	Cys	Gln	Arg	Gly 1205	Tyr		Gly		Trp 1210	_	_	Asp	_	Ile 1215	
Gln	Thr	Thr			Cys		Ala				Gly			Lys	
Cys	Ser	Met 1235	Arg				Pro 1240	Arg		Cys		Asn 1245	Thr		His
Gly	ጥኮጒ			-1 .								_			
				TTE	Asn				Thr	Gly		_	Thr	Pro	Ser
	1250)				1255	5			_	1260)			
	1250 Ala)				1255 Arg		Leu		_	1260 Val)			Glu
Pro 1265	1250 Ala 5) Pro	Asn	Tyr Thr	Ser 1270 Arg	1255 Arg) Val	Ala Gly	Leu Asp	Trp Phe	Arg 1275 His	1260 Val Tyr	Ala	Ala	Glu Gly	Glu 1280 Met
Pro 1265 Tyr	1250 Ala S Val	Pro Glu	Asn Val	Tyr Thr 1285	Ser 1270 Arg	1255 Arg) Val	Ala Gly	Leu Asp	Trp Phe 1290	Arg 1275 His	1260 Val Tyr	Ala Val	Ala Thr	Glu Gly 1295	Glu 1280 Met
Pro 1265 Tyr	1250 Ala S Val	Pro Glu	Asn Val	Tyr Thr 1285 Val	Ser 1270 Arg	1255 Arg) Val	Ala Gly	Leu Asp	Trp Phe 1290 Gln	Arg 1275 His	1260 Val Tyr	Ala Val	Ala Thr	Glu Gly 1295 Glu	Glu 1280 Met
Pro 1265 Tyr Thr	1250 Ala Val Thr	Pro Glu Asp Glu	Asn Val Asn 1300 Val	Tyr Thr 1285 Val Asp	Ser 1270 Arg Lys	1255 Arg) Val Cys	Ala Gly Pro	Leu Asp Cys 1305 Leu	Trp Phe 1290 Gln His	Arg 1275 His Val	1260 Val Tyr Pro	Ala Val Ala	Ala Thr Pro 1310 Pro	Glu Gly 1295 Glu	Glu 1280 Met Dhe
Pro 1265 Tyr Thr	1250 Ala Val Thr	Pro Glu Asp Glu 1319 Leu	Asn Val Asn 1300 Val	Tyr Thr 1285 Val Asp	Ser 1270 Arg Lys Gly	1255 Arg Val Cys Val	Ala Gly Pro Arg 1320 Val	Leu Asp Cys 1305 Leu	Trp Phe 1290 Gln His	Arg 1275 His Val	1260 Val Tyr Pro Tyr Val	Ala Val Ala Ala 1325 Gly	Ala Thr Pro 1310 Pro	Glu Gly 1295 Glu) Ala	Glu 1280 Met Phe Cys
Pro 1265 Tyr Thr Phe Lys	1250 Ala Val Thr Thr Pro 1330	Pro Glu Asp Glu 1319 Leu	Asn Val Asn 1300 Val Leu	Tyr Thr 1285 Val Asp Arg	Ser 1270 Arg Lys Gly	1255 Arg Val Cys Val Glu 1335	Ala Gly Pro Arg 1320 Val	Leu Asp Cys 1305 Leu Thr	Trp Phe 1290 Gln His	Arg 1275 His Val Arg	1260 Val Tyr Pro Tyr Val 1340	Ala Val Ala Ala 1325 Gly	Ala Thr Pro 1310 Pro Leu	Glu 1295 Glu Ala Asn	Glu 1280 Met Phe Cys
Pro 1265 Tyr Thr Phe	1250 Ala Val Thr Thr Pro 1330 Leu	Pro Glu Asp Glu 1319 Leu	Asn Val Asn 1300 Val Leu	Tyr Thr 1285 Val Asp Arg	Ser 1270 Arg Lys Gly	1255 Arg Val Cys Val Glu 1335 Leu	Ala Gly Pro Arg 1320 Val	Leu Asp Cys 1305 Leu Thr	Trp Phe 1290 Gln His	Arg 1275 His Val Arg	1260 Val Tyr Pro Tyr Val 1340 Glu	Ala Val Ala Ala 1325 Gly	Ala Thr Pro 1310 Pro Leu	Glu 1295 Glu Ala Asn	Glu 1280 Met Phe Cys
Pro 1265 Tyr Thr Phe Lys Tyr 1345	1250 Ala Val Thr Thr Pro 1330 Leu	Pro Glu Asp Glu 1319 Leu Val	Asn Val Asn 1300 Val Leu Gly	Tyr Thr 1285 Val Asp Arg Ser Met	Ser 1270 Arg Lys Gly Glu Gln 1350 Leu	1255 Arg Val Cys Val Glu 1335 Leu	Ala Gly Pro Arg 1320 Val	Leu Asp Cys 1305 Leu Thr	Trp Phe 1290 Gln His Phe Glu Ser	Arg 1275 His Val Val Arg Leu Pro 1355 His	1260 Val Tyr Pro Tyr Val 1340 Glu	Ala Val Ala Ala 1325 Gly Pro	Ala Thr Pro 1310 Pro Leu Asp	Glu Gly 1295 Glu Ala Asn Val Glu	Glu 1280 Met Phe Cys Gln Ala 1360 Thr
Pro 1265 Tyr Thr Phe Lys Tyr 1345 Val	1250 Ala Val Thr Thr Pro 1330 Leu	Pro Glu Asp Glu 1315 Leu Val Thr	Asn Val Asn 1300 Val Leu Gly Ser Arg	Tyr Thr 1285 Val Asp Arg Ser Met 1365 Leu	Ser 1270 Arg Lys Gly Glu Gln 1350 Leu	1255 Arg Val Cys Val Glu 1335 Leu Thr	Ala Gly Pro Arg 1320 Val Pro	Leu Asp Cys 1305 Leu Thr Cys Pro	Trp Phe 1290 Gln His Phe Glu Ser 1370 Pro	Arg 1275 His Val Arg Leu Pro 1355 His	1260 Val Tyr Pro Tyr Val 1340 Glu	Ala Val Ala Ala 1325 Gly Pro Thr	Ala Thr Pro 1310 Pro Leu Asp Ala Ala	Glu Gly 1295 Glu Ala Asn Val Glu 1375 Ser	Glu 1280 Met Phe Cys Gln Ala 1360 Thr
Pro 1265 Tyr Thr Phe Lys Tyr 1345 Val	1250 Ala Val Thr Pro 1330 Leu Lys	Pro Glu Asp Glu 1319 Leu Val Thr Arg	Asn Val Asn 1300 Val Leu Gly Ser Arg 1380	Tyr Thr 1285 Val Asp Arg Ser Met 1365 Leu	Ser 1270 Arg Lys Gly Glu Gln 1350 Leu	1255 Arg Val Cys Val Glu 1335 Leu Thr	Ala Gly Pro Arg 1320 Val Pro Asp Gly	Leu Asp Cys 1305 Leu Thr Cys Pro Ser 1385	Trp Phe 1290 Gln His Phe Glu Ser 1370 Pro	Arg 1275 His Val Arg Leu Pro 1355 His	Tyr Pro Tyr Val Glu Ile Ser	Ala Val Ala Ala 1325 Gly Pro Thr	Ala Thr Pro 1310 Pro Leu Asp Ala Ala 1390	Glu Gly 1295 Glu Ala Asn Val Glu 1375 Ser	Glu 1280 Met Phe Cys Gln Ala 1360 Thr
Pro 1265 Tyr Thr Phe Lys Tyr 1345 Val	1250 Ala Val Thr Pro 1330 Leu Lys	Pro Glu Asp Glu 1319 Leu Val Thr Arg	Asn Val Asn 1300 Val Leu Gly Ser Arg 1380 Gln	Tyr Thr 1285 Val Asp Arg Ser Met 1365 Leu	Ser 1270 Arg Lys Gly Glu Gln 1350 Leu	1255 Arg Val Cys Val Glu 1335 Leu Thr	Ala Gly Pro Arg 1320 Val Pro Asp Gly Pro	Leu Asp Cys 1305 Leu Thr Cys Pro Ser 1385 Ser	Trp Phe 1290 Gln His Phe Glu Ser 1370 Pro	Arg 1275 His Val Arg Leu Pro 1355 His	Tyr Pro Tyr Val Glu Ile Ser	Ala Val Ala 1325 Gly Pro Thr Leu Thr	Ala Thr Pro 1310 Pro Leu Asp Ala Ala 1390 Cys	Glu Gly 1295 Glu Ala Asn Val Glu 1375 Ser	Glu 1280 Met Phe Cys Gln Ala 1360 Thr
Pro 1265 Tyr Thr Phe Lys Tyr 1345 Val Ala Ser	1250 Ala Val Thr Pro 1330 Leu Lys Ala His	Pro Glu Asp Glu 1319 Leu Val Thr Arg Ser 1399 Asp	Asn Val Asn 1300 Val Leu Gly Ser Arg 1380 Gln	Tyr Thr 1285 Val Asp Arg Ser Met 1365 Leu Leu	Ser 1270 Arg Lys Gly Glu Gln 1350 Leu Ala	l255 Arg Val Cys Val Glu 1335 Leu Thr Arg Ala	Ala Gly Pro Arg 1320 Val Pro Asp Gly Pro 1400 Asp	Leu Asp Cys 1305 Leu Thr Cys Pro Ser 1385 Ser	Trp Phe 1290 Gln His Phe Glu Ser 1370 Pro	Arg 1275 His Val Arg Leu Pro 1355 His Pro	Tyr Pro Tyr Val 1340 Glu Ile Ser Alá Ala	Ala Val Ala Ala 1325 Gly Pro Thr Leu Thr 1405 Asn	Ala Thr Pro 1310 Pro Leu Asp Ala Ala 1390 Cys	Glu Gly 1295 Glu Ala Asn Val Glu 1375 Ser	Glu 1280 Met Phe Cys Gln Ala 1360 Thr Ser
Pro 1265 Tyr Thr Phe Lys Tyr 1345 Val Ala Ser Arg	1250 Ala Val Thr Pro 1330 Leu Lys Ala His 1410	Pro Glu Asp Glu 1319 Leu Val Thr Arg Ser 1399 Asp	Asn Val Asn 1300 Val Leu Gly Ser Arg 1380 Gln Ser	Tyr Thr 1285 Val Asp Arg Ser Met 1365 Leu Leu Pro	Ser 1270 Arg Lys Gly Glu Gln 1350 Leu Ala Ser	l255 Arg Val Cys Val Glu 1335 Leu Thr Arg Ala Ala 1415	Ala Gly Pro Arg 1320 Val Pro Asp Gly Pro 1400 Asp	Leu Asp Cys 1305 Leu Thr Cys Pro Ser 1385 Ser	Trp Phe 1290 Gln His Phe Glu Ser 1370 Pro Leu Ile	Arg 1275 His Val Arg Leu Pro 1355 His Pro Lys Glu	Tyr Pro Tyr Val 1340 Glu Ile Ser Ala Ala 1420	Ala Val Ala Ala 1325 Gly Pro Thr Leu Thr 1405 Asn	Ala Thr Pro 1310 Pro Leu Asp Ala Ala 1390 Cys Leu	Glu Gly 1295 Glu Ala Asn Val Glu 1375 Ser Thr	Glu 1280 Met Phe Cys Gln Ala 1360 Thr Ser Thr
Pro 1265 Tyr Thr Phe Lys Tyr 1345 Val Ala Ser Arg	1250 Ala Val Thr Pro 1330 Leu Lys Ala His 1410 Gln	Pro Glu Asp Glu 1319 Leu Val Thr Arg Ser 1399 Asp	Asn Val Asn 1300 Val Leu Gly Ser Arg 1380 Gln Ser	Tyr Thr 1285 Val Asp Arg Ser Met 1365 Leu Leu Pro	Ser 1270 Arg Lys Gly Glu Gln 1350 Leu Ala Ser	l255 Arg Val Cys Val Glu 1335 Leu Thr Arg Ala 1415 Asn	Ala Gly Pro Arg 1320 Val Pro Asp Gly Pro 1400 Asp	Leu Asp Cys 1305 Leu Thr Cys Pro Ser 1385 Ser	Trp Phe 1290 Gln His Phe Glu Ser 1370 Pro Leu Ile	Arg 1275 His Val Arg Leu Pro 1355 His Pro Lys	Tyr Pro Tyr Val 1340 Glu Ile Ser Ala Ala 1420 Glu	Ala Val Ala Ala 1325 Gly Pro Thr Leu Thr 1405 Asn	Ala Thr Pro 1310 Pro Leu Asp Ala Ala 1390 Cys Leu	Glu Gly 1295 Glu Ala Asn Val Glu 1375 Ser Thr	Glu 1280 Met Phe Cys Gln Ala 1360 Thr Ser Thr Trp Lys
Pro 1265 Tyr Thr Phe Lys Tyr 1345 Val Ala Ser Arg 1425	1250 Ala Val Thr Pro 1330 Leu Lys Ala His 1410 Gln	Pro Glu Asp Glu 1319 Leu Val Thr Arg Ser 1399 Asp Glu	Asn Val Asn 1300 Val Leu Gly Ser Arg 1380 Gln Ser Met	Tyr Thr 1285 Val Asp Arg Ser Met 1365 Leu Pro Gly Asp	Ser 1270 Arg Lys Gly Glu Gln 1350 Leu Ala Ser Asp Gly 1430 Ser	l255 Arg Val Cys Val Glu 1335 Leu Thr Arg Ala 1415 Asn	Ala Gly Pro Arg 1320 Val Pro Asp Gly Pro 1400 Asp	Leu Asp Cys 1305 Leu Thr Cys Pro Ser 1385 Ser Leu Thr	Trp Phe 1290 Gln His Phe Glu Ser 1370 Pro Leu Ile Arg	Arg 1275 His Val Arg Leu Pro 1355 His Pro Lys Glu Val 1435 Gln	Tyr Pro Tyr Val 1340 Glu Ile Ser Ala Ala 1420 Glu	Ala Val Ala 1325 Gly Pro Thr Leu Thr 1405 Asn	Ala Thr Pro 1310 Pro Leu Asp Ala 1390 Cys Leu Glu	Glu Gly 1295 Glu Ala Asn Val Glu 1375 Ser Thr Leu Asn Asp	Glu 1280 Met Phe Cys Gln Ala 1360 Thr Ser Thr Trp Lys 1440 Glu
Pro 1265 Tyr Thr Phe Lys Tyr 1345 Val Ala Ser Arg 1425 Val	Thr Pro 1330 Leu Lys Ala His 1410 Gln Val	Pro Glu Asp Glu 1319 Leu Val Thr Arg Ser 1399 Asp Glu Ile	Asn Val Asn 1300 Val Leu Gly Ser Arg 1380 Gln Ser Met	Tyr Thr 1285 Val Asp Arg Ser Met 1365 Leu Pro Gly Asp 1445	Ser 1270 Arg Lys Gly Glu Gln 1350 Leu Ala Ser Asp Gly 1430 Ser	Val Cys Val Glu 1335 Leu Thr Arg Ala 1415 Asn	Ala Gly Pro Arg 1320 Val Pro Asp Gly Pro 1400 Asp Ile	Leu Asp Cys 1305 Leu Thr Cys Pro Ser 1385 Ser Leu Thr	Trp Phe 1290 Gln His Phe Glu Ser 1370 Pro Leu Ile Arg Leu 1450	Arg 1275 His Val Arg Leu Pro 1355 His Pro Lys Glu Val 1435 Gln	Tyr Pro Tyr Val 1340 Glu Ile Ser Ala Ala 1420 Glu Ala	Ala Val Ala Ala 1325 Gly Pro Thr Leu Thr 1405 Asn Ser Glu	Ala Thr Pro 1310 Pro Leu Asp Ala 1390 Cys Leu Glu Glu	Glu Gly 1295 Glu Ala Asn Val Glu 1375 Ser Thr Leu Asn Asp 1455	Glu 1280 Met Phe Cys Gln Ala 1360 Thr Ser Thr Lys 1440 Glu

26 / 93

Pro Arg Ala Met Pro Ile Trp Ala Arg Pro Asp Tyr Asn Pro Pro Leu Xaa Glu Ser Trp Lys Asp Pro Asp Tyr Val Pro Pro Val Val His Gly Cys Pro Leu Pro Pro Ala Lys Ala Pro Pro Ile Pro Pro Pro Arg Arg Lys Arg Thr Val Val Leu Ser Glu Ser Thr Val Ser Ser Ala Leu Ala Glu Leu Ala Thr Lys Thr Phe Gly Ser Ser Glu Ser Ser Ala Val Asp Ser Gly Thr Ala Thr Ala Ser Pro Asp Gln Pro Ser Asp Asp Gly Asp Ala Gly Ser Asp Val Glu Ser Tyr Ser Ser Met Pro Pro Leu Glu Gly Glu Pro Gly Asp Pro Asp Leu Ser Asp Gly Ser Trp Ser Thr Val Ser Glu Glu Ala Ser Glu Asp Val Val Cys Cys Ser Met Ser Tyr Thr Trp Thr Gly Ala Leu Ile Thr Pro Cys Ala Ala Glu Glu Thr Lys Leu Pro Ile Asn Ala Leu Ser Asn Ser Leu Leu Arg His His Asn Leu Val Tyr Ala Thr Thr Ser Arg Ser Ala Ser Leu Arg Gln Lys Lys Val Thr Phe 1660. Asp Arg Leu Gln Val Leu Asp Asp His Tyr Arg Asp Val Leu Lys Glu Met Lys Ala Lys Ala Ser Thr Val Lys Ala Lys Leu Leu Ser Val Glu Glu Ala Cys Lys Leu Thr Pro Pro His Ser Ala Arg Ser Lys Phe Gly Tyr Gly Ala Lys Asp Val Arg Asn Leu Ser Ser Lys Ala Val Asn His Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr Cys Tyr Leu Lys Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys Glu Ser

```
Ala Gly Thr Glu Asp Glu Ala Ser Leu Arg Ala Phe Thr Glu Ala
            1940
                                1945
Met Thr Arg Tyr Ser Ala Pro Pro Gly Asp Pro Pro Lys Pro Glu Tyr
                            1960
Asp Leu Glu Leu Ile Thr Ser Cys Ser Ser Asn Val Ser Val Ala His
    1970
                        1975
                                             1980
Asp Ala Ser Gly Lys Arg Val Tyr Tyr Leu Thr Arg Asp Pro Thr Thr
1985
                   1990
                                         1995
                                                             2000
Pro Leu Ala Arg Ala Ala Trp Glu Thr Ala Arg His Thr Pro Val Asn
                2005
                                     2010
                                                         2015
Ser Trp Leu Gly Asn Ile Ile Met Tyr Ala Pro Thr Leu Trp Ala Arg
            2020
                                2025
Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln
        2035
                            2040
                                                2045
Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile
    2050
                        2055
                                            2060
Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser
2065
                    2070
                                        2075
                                                             2080
Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala
                2085
                                    2090
Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His
            2100
                                2105
                                                    2110
Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala
        2115
                            2120
                                                2125
Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu
    2130
                        2135
                                            2140
Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp
2145
                    2150
                                        2155
Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg
                2165
                                    2170
                                                         2175
Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly
            2180
                                2185
                                                    2190
Val Gly Ile Tyr Leu Leu Pro Asn Arg
        2195
                            2200
<210> 4
<211> 8643
<212> DNA
<213> HCV
<220>
<221> CDS
<222> (1802)...(8407)
<400> 4
accagecece gattggggge gacaétecae catagateae teceetgtga ggaactaetg 60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgtcgtgcag cctccaggac 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg teetttettg gateaacccg eteaatgeet ggagatttgg gegtgeecc 240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc gcaacagaca atcggctgct 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtatccat catggctgat gcaatgcggc ggctgcatac gcttgatccg gctacctgcc 780
```

ttgtcgatca ccaggctcaa gcttgccgaa tgggtgtggc ttggcggcga agcgcatcgc cagaccacaa ccgaagccgc gccgtctttt taggggtctt taggggtctt agttcctctg gaaccccca tgcaaaggcg atggctctcc tatgggatct aaacgtctag c atg gac	ggatgatctg ggcgcgcatg tatcatggtg ggaccgctat atgggctgac cttctatcgc cggtttccct ttggaataag ggcaatgtga tccctctcg gaagcttctt cctggcgaca gcacaacccc tcaagcgtat gatctggggc gcccccgaa gg gag atg	gacgaagagagagagagagagagagagagagagagagag	atcaggggci aggatctcgi gcttttctgi gcgttggctag gagttcttctgi agttcttctgi caattccgci acctggccci acctggccci gcaaggtctgi acggcaaaag gctgaaggai gctgaaggai gctgaaggai gtggtttci gtggtttci	tactcggatg cgcgccagcc cgtgacccat attcatcgac ccgtgatatt tatcgccgct agttcgcgcc cccccccta atgttattt gtcttcttga ttgaatgtcg ccacgtgtat gccacgaagg tgtgtttagt tttgaaaaac gc gcg gtt t	gaactgttcg ggcgatgcct tgtggccggc gctgaagagc ccgattcgc cagatgttaa acgttactgg ccaccatatt cgagcattcc tgaaggaagc gcaggcagcg aagatacacc aaagagtcaa taccccattg cgaggttaga acgataatac tc gta ggt	900 960 1020 1080 1140 1200 1260 1320 1380 1440 1500 1560 1620 1680 1740 1800
				g ctg ttc ctc s Leu Phe Leu 30	Ala Arg	1897
	Trp Leu G			g gcc gag gca g Ala Glu Ala 45	_	1945
				g ggc cgc gat y Gly Arg Asp 60 .		1993
		_	~ ~	a atc ttt acc 1 Ile Phe Thr 5		2041
				g gtg ctc cag t Val Leu Gln	7 00	2089
			_	c ggg ctc att s Gly Leu Ile 110	Arg Ala	2137
-	Val Arg I		_	t tat gtc caa s Tyr Val Gln 125		2185
		-		c gtt tat gac r Val Tyr Asp 140	_	2233
	Arg Asp 1	- -		a cga gac ctt u Arg Asp Leu 5		2281

									gtt Val		acc Thr	2329
tgg Trp					=			_			ccc Pro	2377
gtc Val								 =		_	agc Ser	2425
Leu	_	_	 _								tcc Ser	2473
caa Gln 225								-			ggc Gly 240	2521
cgg Arg											gca Ala	2569
aca Thr												2617
tat Tyr												2665
acc Thr												2713
ecc Pro 305												2761
ctt Leu				Lys	Ala	Asp	Ile	Val	Arg			2809
ggc								_			_	2857
aag Lys												2905
Gly												2953

	ttt Phe	_		 				_	_			_	_	3001
_	acg Thr	_			_	_		_			_		gtg Val	3049
	cat His												ccg Pro	3097
_	gcg Ala		_						_			=		3145
	gcc Ala 450												ggt Gly	3193
	gac Asp													3241
	atc Ile												tgc Cys	3289
	ely aaa												act Thr	3337
	tcg Ser												gag Glu	3385
_	gct Ala 530				 		•			_		_	gga Gly	3433
	gtc Val												agc Ser 560	3481
	gga Gly													3529
	Gly 999													3577
- - -	ctc Leu			_						_	_	_		3625

		_			gta Val		_	_								3673
					gct Ala 630											3721
					aat Asn								_		agc Ser	3769
				Phe	acc Thr										gcg Ala	3817
					cgg Arg										Gly	3865
att Ile	tac Tyr 690	agg Arg	ttt Phe	gtg Val	act Thr	cca Pro 695	gga Gly	gaa Glu	cgg Arg	ccc Pro	tcg Ser 700	ggc	atg Met	ttc Phe	gat Asp	3913
					gag Glu 710										gag Glu 720	3961
					acc Thr										aca Thr	4009
					tgc Cys										gtc Val	4057
					cac His						_		_		aag Lys	4105
					ttc Phe										gtg Val	4153
					gct Ala 790				_	-			_		_	4201
		_			aag Lys										- -	4249
				_	gtt Val											4297

	atc atg gca Ile Met Ala		_		
	gtg ctg gta Val Leu Val				_
_	aca ggc agc Thr Gly Ser 870				
	gcc atc att Ala Ile Ile 885			Tyr Arg	
-	gaa gag tgc Glu Glu Cys 900	_			
	gcc gaa caa Ala Glu Gln				
——————————————————————————————————————	aag caa gcg Lys Gln Ala			Val Glu S	-
	ctc gaa gcc Leu Glu Ala 950				_
	caa tat tta Gln Tyr Leu 965	-	-	Pro Gly A	
	tca ctg atg Ser Leu Met 980		-	_	-
	cat acc ctc His Thr Leu 5			-	
_	gct cct ccc Ala Pro Pro		-	Val Gly A	-
	gcg gct gtt Ala Ala Val 1030	Gly Ser Ile	-	= = =	
	gca ggt tat Ala Gly Tyr 1045			Ala Leu V	

_	er Gly Glu Met P	cc tcc acc gag gac ro Ser Thr Glu Asp .065	
		gc gcc cta gtc gtc ly Ala Leu Val Val 1085	Gly Val Val
		tg ggc cca ggg gag al Gly Pro Gly Glu 1100	
_		tc gct tcg cgg ggt he Ala Ser Arg Gly 1115	
Ser Pro Thr His T		gc gac gct gca gca er Asp Ala Ala Ala 1130	
	er Leu Thr Ile T	ct cag ctg ctg aag hr Gln Leu Leu Lys 145	
	_	cg cca tgc tcc ggc hr Pro Cys Ser Gly 1165	Ser Trp Leu
_	_	cg gtg ttg act gat hr Val Leu Thr Asp 1180	
		ga ttg ccg gga gtc rg Leu Pro Gly Val 1195	
Ser Cys Gln Arg G		tc tgg cgg ggc gac al Trp Arg Gly Asp 1210	
Gln Thr Thr Cys P:	co Cys Gly Ala G	ag atc acc gga cat In Ile Thr Gly His 225	
		gg acc tgt agt aac rg Thr Cys Ser Asn 1245	Thr Trp His
		cc acg ggc ccc tgc hr Thr Gly Pro Cys 1260	
		tg tgg cgg gtg gct eu Trp Arg Val Ala 1275	_

tac gtg gag gtt ac Tyr Val Glu Val Th	r Arg Val Gly		-	
acc act gac aac gt Thr Thr Asp Asn Va 1300	l Lys Cys Pro	_		Glu Phe
ttc aca gaa gtg ga Phe Thr Glu Val As 1315	· ·	Leu His Arg	_	
aaa ccc ctc cta cg Lys Pro Leu Leu Ar 1330				
tac ctg gtt ggg tc Tyr Leu Val Gly Se 1345			Glu Leu Asp	
gtg ctc act tcc at Val Leu Thr Ser Me 13	Leu Thr Asp			
gct aag cgt agg ct Ala Lys Arg Arg Le 1380	Ala Arg Gly			Ser Ser
tca gct agc cag ct Ser Ala Ser Gln Le 1395		Ser Leu Lys		
cgt cat gac tcc cc Arg His Asp Ser Pr 1410				_ _
cgg cag gag atg gg Arg Gln Glu Met Gl 1425			Glu Ser Glu	-
gta gta att ttg ga Val Val Ile Leu As 14		Pro Leu Gln		
agg gaa gta tcc gt		atc ctg cgg	agg tcc agg	
Arg Glu Val Ser Va 1460				
	ata tgg gca	Ile Leu Arg 1465 cgc ccg gat Arg Pro Asp	Arg Ser Arg 1470 tac aac cct	cca ctg 6265

	Pro				gcc Ala 151	Lys					Pro				agg Arg 1520	6361
					Leu					Val						6409
gag Glu	ctc Leu	gcc Ala	aca Thr 1540	Lys	acc Thr	ttc Phe	ggc Gly	agc Ser 154	Ser	gaa Glu	tcg Ser	tcg Ser	gcc Ala 155	Val	gac Asp	6457
agc Ser	ggć Gly	acg Thr 155!	Ala	acg Thr	gcc Ala	tct Ser	ect Pro 1560	Asp	cag Gln	ccc	tcc Ser	gac Asp 156!	Asp	ggc Gly	gac Asp	6505
		Ser			gag Glu		Tyr					Pro			Gly 999	6553
gag Glu 1585	Pro	gjå aaa	gat Asp	ccc Pro	gat Asp 1590	Leu	agc Ser	gac Asp	gjå aaa	tct Ser 1599	Trp	tct Ser	aċc Thr	gta Val	agc Ser 1600	6601
Glu	Glu	Ala	Ser	Glu 1605		Val	Val	Cys	Cys 1610	Ser	Met	Ser	Tyr	Thr 1615	Trp	6649
acg Thr	Gly	gcc Ala	ctg Leu 1620	Ile	acg Thr	cca Pro	tgc Cys	gct Ala 1625	Ala	gag Glu	gaa Glu	acc Thr	aag Lys 1630	Leu	ccc Pro	6697
			Leu		aac Asn			Leu					Leu			6745
gct Ala	aca Thr 1650	Thr	tct Ser	cgc Arg	agc Ser	gca Ala 1655	Ser	ctg Leu	cgg Arg	cag Gln	aag Lys 1660	Lys	gtc Val	acc Thr	ttt Phe	6793
gac Asp 1665	Arg	ctg Leu	cag Gln	gtc Val	ctg Leu 1670	Asp	Asp	cac His	Tyr	Arg	Asp	gtg Val	Leu	Lys	gag Glu 1680	6841
atg Met	aag Lys	gcg Ala	aag Lys	gcg Ala 1685	tcc Ser	aca Thr	gtt Val	aag Lys	gct Ala 1690	Lys	ctt Leu	cta Leu	tcc Ser	gtg Val 1695	Glu	6889
				Leu	acg Thr				Ser					Phe		6937
			Lys		gtc Val			Leu					Val			6985

atc cgc tcc gtg tgg aag gac ttg ctg gaa gac act gag aca cca att Ile Arg Ser Val Trp Lys Asp Leu Leu Glu Asp Thr Glu Thr Pro Ile 1730 1735 1740	7033
gac acc acc atc atg gca aaa aat gag gtt ttc tgc gtc caa cca gag Asp Thr Thr Ile Met Ala Lys Asn Glu Val Phe Cys Val Gln Pro Glu 1745 1750 1760	7081
aag ggg ggc cgc aag cca gct cgc ctt atc gta ttc cca gat ttg ggg Lys Gly Gly Arg Lys Pro Ala Arg Leu Ile Val Phe Pro Asp Leu Gly 1765 1770 1775	7129
gtt cgt gtg tgc gag aaa atg gcc ctt tac gat gtg gtc tcc acc ctc Val Arg Val Cys Glu Lys Met Ala Leu Tyr Asp Val Val Ser Thr Leu 1780 1785 . 1790	7177
cct cag gcc gtg atg ggc tct tca tac gga ttc caa tac tct cct gga Pro Gln Ala Val Met Gly Ser Ser Tyr Gly Phe Gln Tyr Ser Pro Gly 1795 1800 1805	7225
cag cgg gtc gag ttc ctg gtg aat gcc tgg aaa gcg aag aaa tgc cct Gln Arg Val Glu Phe Leu Val Asn Ala Trp Lys Ala Lys Lys Cys Pro 1810 1815 1820	7273
atg ggc ttc gca tat gac acc cgc tgt ttt gac tca acg gtc act gag Met Gly Phe Ala Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu 1825 1830 1835 1840	7321
aat gac atc cgt gtt gag gag tca atc tac caa tgt tgt gac ttg gcc Asn Asp Ile Arg Val Glu Glu Ser Ile Tyr Gln Cys Cys Asp Leu Ala 1845 1850 1855	7369
ccc gaa gcc aga cag gcc ata agg tcg ctc aca gag cgg ctt tac atc Pro Glu Ala Arg Gln Ala Ile Arg Ser Leu Thr Glu Arg Leu Tyr Ile 1860 1865 1870	7417
ggg ggc ccc ctg act aat tct aaa ggg cag aac tgc ggc tat cgc cgg Gly Gly Pro Leu Thr Asn Ser Lys Gly Gln Asn Cys Gly Tyr Arg Arg 1875 1880 1885	7465
tgc cgc gcg agc ggt gta ctg acg acc agc tgc ggt aat acc ctc aca	7513
Cys Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr Leu Thr 1890 1895 1900	
	7561
tgt tac ttg aag gcc gct gcg gcc tgt cga gct gcg aag ctc cag gac Cys Tyr Leu Lys Ala Ala Ala Ala Cys Arg Ala Ala Lys Leu Gln Asp	7561 7609

Met Thr Arg Tyr 1955			ccg ccc aaa e Pro Pro Lys 1 1965		705
gac ttg gag ttg Asp Leu Glu Leu 1970		Cys Ser Ser			753
gat gca tct ggc Asp Ala Ser Gly 1985			- -		7801
ccc ctt gcg cgg Pro Leu Ala Arg			Arg His Thr		7849
tcc tgg cta ggc Ser Trp Leu Gly 2020	Asn Ile Ile		Pro Thr Leu		7897
atg atc ctg atg Met Ile Leu Met 2035					7945
ctt gaa aaa gcc Leu Glu Lys Ala 2050		Gln Ile Tyr	-		7993
gag cca ctt gac Glu Pro Leu Asp 2065					3041
gca ttt tca ctc Ala Phe Ser Leu			Glu Ile Asn	•	3089
	His Ser Tyr 2085 aaa ctt ggg Lys Leu Gly	Ser Pro Gly 209 gta ccg ccc	Glu Ile Asn A O ttg cga gtc t Leu Arg Val	Arg Val Ala 2095 tgg aga cat 8	3089
tca tgc ctc agg Ser Cys Leu Arg 2100 cgg gcc aga agt Arg Ala Arg Ser	His Ser Tyr 2085 aaa ctt ggg Lys Leu Gly gtc cgc gct	Ser Pro Gly 209 gta ccg ccc Val Pro Pro 2105	ttg cga gtc the Leu Arg Val tcc cag ggg g	Arg Val Ala 2095 tgg aga cat 8 Trp Arg His 2110	
tca tgc ctc agg Ser Cys Leu Arg 2100 cgg gcc aga agt Arg Ala Arg Ser	His Ser Tyr 2085 aaa ctt ggg Lys Leu Gly gtc cgc gct Val Arg Ala aag tac ctc	Ser Pro Gly 209 gta ccg ccc Val Pro Pro 2105 agg cta ctg Arg Leu Leu 2120 ttc aac tgg Phe Asn Trp	ttg cga gtc the Leu Arg Val 3 tcc cag ggg 9 Ser Gln Gly 9 2125	Arg Val Ala 2095 egg aga cat 8 frp Arg His 2110 egg agg gct 8 fly Arg Ala acc aag ctc 8	3137
tca tgc ctc agg Ser Cys Leu Arg 2106 cgg gcc aga agt Arg Ala Arg Ser 2115 gcc act tgt ggc Ala Thr Cys Gly	His Ser Tyr 2085 aaa ctt ggg Lys Leu Gly gtc cgc gct Val Arg Ala aag tac ctc Lys Tyr Leu 213!	Ser Pro Gly 209 gta ccg ccc Val Pro Pro 2105 agg cta ctg Arg Leu Leu 2120 ttc aac tgg Phe Asn Trp 5	ttg cga gtc the Leu Arg Val 3 tcc cag ggg 9 Ser Gln Gly 0 2125 gca gta agg 8 Ala Val Arg 2 2140 ttg gat tta 1	Arg Val Ala 2095 agg aga cat Erp Arg His 2110 agg agg gct Bly Arg Ala acc aag ctc Ehr Lys Leu tcc agc tgg 8	3137 3185

38 / 93

gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg 8377 Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly 2190 2180 2185 8427 gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca Val Gly Ile Tyr Leu Leu Pro Asn Arg * 2195 2200 ttttttttt ttttttt ttttctttt tcccaatttt tttcctttt tttcctttgg 8547 tggctccatc ttagccctag tcacggctag ctgtgaaagg tccgtgagcc gcttgactgc 8607 agagagtgct gatactggcc tctctgcaga tcaagt 8643 <210> 5 <214> 8648 <212> DNA <213> HCV <220> <221> CDS <222> (1802)...(8407) <400> 5 gccagccccc gattgggggc gacactccac catagatcac tcccctgtga ggaactactg 60 tetteacgea gaaagegtet agecatggeg ttagtatgag tgtegtgeag cetecaggae 120 ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180 gacgaccggg teettettg gatcaacccg etcaatgeet ggagatttgg gegtgeecee 240 gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300 gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360 ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420 eggeegettg ggtggagagg etattegget atgaetggge acaacagaea ateggetget 480 ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540 acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600 cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660 tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720 aagtateeat catggetgat geaatgegge ggetgeatae gettgateeg getaeetgee 780 cattegacea ecaagegaaa categeateg agegageaeg tacteggatg gaageeggte 840 ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900 ccaggeteaa ggegegeatg eccgaeggeg aggatetegt egtgaeceat ggegatgeet 960 gcttgccgaa tatcatggtg gaaaatggcc gcttttctgg attcatcgac tgtggccggc 1020 tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080 ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc 1140 agegeatege ettetatege ettettgaeg agttettetg agttegegee eagatgttaa 1200 cagaccacaa cggtttccct ctagcgggat caattccgcc cccccccta acgttactgg 1260 ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320 gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380 taggggtctt tcccctctcg ccaaaggaat gcaaggtctg ttgaatgtcg tgaaggaagc 1440 agttectetg gaagettett gaagacaaac aacgtetgta gegaceettt geaggeageg 1500 gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560 tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620 atggctctcc tcaagcgtat tcaacaaggg gctgaaggat gcccagaagg taccccattg 1680 tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740 aaacgtctag gcccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800 c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849 Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly 10 15

	ata Ile		_			_	_						_	agg Arg	1897
	ata Ile									=		_		ttg Leu	1945
	gtg Val 50												_	gtc Val	1993
	ctc Leu			-							•				2041
	atc Ile					-			_	_		_	_		2089
	acc Thr														2137
	atg Met						·						_	gct Ala	2185
	atg Met 130		_		-					_		_		ctc Leu	2233
	Pro					_		_						gtg Val 160	2281
	gtt Val		_	=			-					_			2329
	GJÀ aaa		Thr	Ala		Cys	Gly	Asp	Ile	Ile	Leu	${ t Gly}$	_	ccc Pro	2377
_	tcc Ser	_						_						agc Ser	2425
	gaa Glu 210				_			_			=			tcc Ser	2473
	cag Gln						_								2521

cgg ga							Glu								2569
aca ca															2617
tat car Tyr Hi		-			•										2665
acc car Thr Gl: 29	n Met		_				_						_	_	2713
ccc ccc Pro Pro 305		_			-										2761
ctt ta Leu Ty	_	- -				-				-					2809
ggc ga Gly As	· -		_								Val				2857
aag gg Lys Gl		- -				_				-	-			-	2905
ggc at Gly Il 37	e Phe			_				_		_	_				2953
gac tt Asp Ph 385			_							_				_	3001
ttc ac			Ser		Pro	Pro	Ala	Val	Pro	Gln	Thr	Phe	Gln	Val	3049
gcc ca Ala Hi			_				-	· _	_	_		_		_	3097
gct gc Ala Al	_	Ala	_				-			_	_		_		3145
gtc gc Val Al 45	a Ala										_	_			3193

	gac Asp														gcc Ala 480	3241
	atc Ile							-				_			tgc Cys	3289
	gly aaa	_	_					_							act Thr	3337
	tcg Ser														gag Glu	3385
_	gct Ala 530			_			_				_	_		_	gga Gly	3433
tcg Ser 545	gtc Val	acc Thr	gtg Val	cca Pro	cat His 550	cca Pro	aac Asn	atc Ile	gag Glu	gag Glu 555	gtg Val	gct Ala	ctg Leu	tcc Ser	agc Ser 560	3481
	gga Gly														atc Ile	3529 ·
	GJÀ aaa															3577
	ctc Leu															3625
	cgg Arg 610												_		att Ile	3673
	gta Val															3721
	gtg Val															3769
	gac Asp															3817
	tca Ser															3865

		agg Arg													gat Asp	3913
		gtt Val			•										gag Glu 720	3961
		ccc Pro													aca Thr	4009
	_	ttg Leu				_		_							gtc Val	4057
		ggc Gly 755									_		_		aag Lys	4105
_		gga Gly								_			_		gtg Val	4153
		agg Arg													aag Lys 800	4201
		ata Ile													ctg Leu	4249
		ctg Leu	Gly	_								•				4297
			820				-	825					830			
		tac Tyr 835	atc					825 tcg	gct	gac	ctg	gag	gtc			4345
Thr	Lys	Tyr 835 tgg Trp	atc Ile gtg	Met ctg Leu	Ala gta Val	Cys ggc Gly	Met 840 gga Gly	tcg Ser gtc Val	gct Ala cta	gac Asp gca Ala	ctg Leu	gag Glu 845 ctg	gtc Val gcc	Val gcg	Thr	4345 4393
agc Ser	acc Thr 850	Tyr 835 tgg Trp	atc Ile gtg Val	Met ctg Leu	Ala gta Val	ggc Gly 855 gtg	Met 840 gga Gly gtc	tcg Ser gtc Val	gct Ala cta Leu	gac Asp gca Ala	ctg Leu gct Ala 860	gag Glu 845 ctg Leu	gtc Val gcc Ala	yal gcg Ala	Thr tat Tyr	
agc Ser tgc Cys 865	acc Thr 850 ctg Leu	Tyr 835 tgg Trp	atc Ile gtg Val aca Thr	Met ctg Leu ggc Gly	Ala gta Val agc Ser 870	ggc Gly 855 gtg Val	Met 840 gga Gly gtc Val	tcg Ser gtc Val att Ile	gct Ala cta Leu gtg Val	gac Asp gca Ala ggc Gly 875	ctg Leu gct Ala 860 agg Arg	gag Glu 845 ctg Leu atc Ile	gtc Val gcc Ala atc Ile	yal gcg Ala ttg Leu	tat Tyr tcc Ser 880	4393

		Ala	gaa caa Glu Gln											4585
			caa gcg Gln Ala											4633
			gaa gcc Glu Ala 950	Phe										4681
Ser	Gly Ile	Gln	tat tta Tyr Leu 965	Ala	Gly	Leu	Ser 970	Thr	Leu	Pro	Gly	Asn 975	Pro	4729
Ala	Ile Ala	Ser 980		Ala	Phe	Thr 985	Ala	Ser	Ile	Thr	Ser 990	Pro	Leu	4777
Thr	Thr Gln 99	His 5	acc ctc Thr Leu	Leu	Phe 1000	Asn)	Ile	Leu	Gly	Gly 1005	Trp	Val	Ala	4825
Ala	Gln Leu 1010	Ala	cct ccc Pro Pro	Ser 1015	Ala	Ala	Ser	Ala	Phe 1020	Val	Gly	Ala	Gly	4873
Ile 1025	Ala Gly	Ala	gct gtt Ala Val 103	Gly O	Ser	Ile	Gly	Leu 1035	Gly	Lys	Val	Leu	Val 1040	4921
Asp	Ile Leu	Ala	ggt tat Gly Tyr 1045	Gly	Ala	Gly	Val 1050	Ala)	Gly	Ala	Leu	Val 1055	Ala	4969
Phe	Lys Val	Met 1060		Glu	Met	Pro 1065	Ser	Thr	Glu	Asp	Leu 1070	Val	Asn	5017
		Ala	atc ctc Ile Leu		Pro	Gly		Leu	Val		Gly			5065
Cys			ctg cgt Leu Arg		His					Glu				5113
_	Trp Met		cgg ctg Arg Leu 111	Ile			_		Arg					5161
		His '	tat gtg Tyr Val 1125				•	Ala					Thr	5209

cag atc ctc tct ag Gln Ile Leu Ser Se 1140		Gln Leu Leu Lys A	
cag tgg atc aac ga Gln Trp Ile Asn Gl 1155			
aga gat gtt tgg ga Arg Asp Val Trp As 1170			•
tgg ctc cag tcc aa Trp Leu Gln Ser Ly 1185		•	_
tca tgt caa cgt gg Ser Cys Gln Arg Gl 12	y Tyr Lys Gly Val		
caa acc acc tgc cc Gln Thr Thr Cys Pr 1220		Ile Thr Gly His V	
tgt tcc atg agg at Cys Ser Met Arg Il 1235			
gga aca ttc ccc at Gly Thr Phe Pro Il 1250	- -		_
ccg gcg cca aat ta Pro Ala Pro Asn Ty 1265		- :	
tac gtg gag gtt ac Tyr Val Glu Val Th 12	r Arg Val Gly Asp	_ :	
acc act gac aac gt Thr Thr Asp Asn Va 1300	_	Gln Val Pro Ala P	
ttc aca gaa gtg ga Phe Thr Glu Val As 1315			- - -
aaa ccc ctc cta cg Lys Pro Leu Leu Ar 1330			
tac ctg gtt ggg tc Tyr Leu Val Gly Se 1345	_		

gtg ctc act tcc at Val Leu Thr Ser Me 13	t Leu Thr Asp P		le Thr Ala Gl	-
gct aag cgt agg ct Ala Lys Arg Arg Le 1380	u Ala Arg Gly S			
tca gct agc cag ct Ser Ala Ser Gln Le 1395		-	-	
cgt cat gac tcc cc Arg His Asp Ser Pr 1410		eu Ile Glu Al		-
cgg cag gag atg gg Arg Gln Glu Met Gl 1425			_	
gta gta att ttg ga Val Val Ile Leu As 14	o Ser Phe Glu P		la Glu Glu As	
agg gaa gta tcc gt Arg Glu Val Ser Va 1460	l Pro Ala Glu I	_	_	
cct cga gcg atg cc Pro Arg Ala Met Pr 1475				
tta gag tcc tgġ aa Leu Glu Ser Trp Ly 1490		yr Val Pro Pr		- - -
tgt cca ttg ccg cc Cys Pro Leu Pro Pr 1505				— — — — — — — — — — — — — — — — — — —
aag agg acg gtt gt Lys Arg Thr Val Va 15	l Leu Ser Glu S		er Ser Ala Le	
gag ctc gcc aca aa Glu Leu Ala Thr Ly 1540	s Thr Phe Gly S			
agc ggc acg gca ac Ser Gly Thr Ala Th 1555				
gcg gga tcc gac gt Ala Gly Ser Asp Va	t gag teg tac t	cc tcc atg cc	c ccc ctt qa	g ggg 6553

gag ccg ggg gat Glu Pro Gly Asp 1585	·			_
gag gag gct agt Glu Glu Ala Ser				Trp
aca ggc gcc ctg Thr Gly Ala Leu 1620	Ile Thr Pro Cys			
atc aat gca ctg Ile Asn Ala Leu 1635		l Leu Arg His H		
gct aca aca tct Ala Thr Thr Ser 1650		Leu Arg Gln L		
gac aga ctg cag Asp Arg Leu Gln 1665				
atg aag gcg aag g Met Lys Ala Lys				Glu
gaa gcc tgt aag Glu Ala Cys Lys 1700	Leu Thr Pro Pro	_		
tat ggg gca aag g Tyr Gly Ala Lys 1 1715		Leu Ser Ser L	- -	
atc cgc tcc gtg 1 Ile Arg Ser Val 1 1730		Leu Glu Asp T		
gac acc acc atc a Asp Thr Thr Ile I 1745	•	Glu Val Phe C		-
aag ggg ggc cgc a Lys Gly Gly Arg		_		Gly
gtt cgt gtg tgc g Val Arg Val Cys (1780	Glu Lys Met Ala		_ _	
cct cag gcc gtg a Pro Gln Ala Val 1 1795		Tyr Gly Phe G		

Gln Arg Val G 1810			tgg aaa gcg Trp Lys Ala 1820	Lys Lys (
atg ggc ttc g Met Gly Phe A 1825		Thr Arg Cys			
aat gac atc c Asn Asp Ile A				Cys Asp I	
ccc gaa gcc a Pro Glu Ala A 1		- -	Leu Thr Glu		
ggg ggc ccc c Gly Gly Pro L 1875					
tgc cgc gcg a Cys Arg Ala S 1890				Asn Thr I	
tgt tac ttg a Cys Tyr Leu L 1905		Ala Ala Cys			
tgc acg atg c Cys Thr Met L					Glu Ser
	1925		1930		1935
gcg ggg acc c Ala Gly Thr G	1925 aa gag gac	gag gcg agc	cta cgg gcc Leu Arg Ala	ttc acg	gag gct 7657
gcg ggg acc c Ala Gly Thr G	1925 aa gag gac ln Glu Asp 940 ac tct gcc yr Ser Ala	gag gcg agc Glu Ala Ser 194! ccc cct ggg	cta cgg gcc Leu Arg Ala	ttc acg con the Thr of 1950	gag gct 7657 Glu Ala gaa tac 7705
gcg ggg acc can all all at act aga tan act act aga tan act act aga tan act	1925 aa gag gac ln Glu Asp 940 ac tct gcc yr Ser Ala tg ata aca eu Ile Thr	gag gcg agc Glu Ala Ser 1949 ccc cct ggg Pro Pro Gly 1960 tca tgc tcc	cta cgg gcc Leu Arg Ala gac ccg ccc Asp Pro Pro tcc aat gtg Ser Asn Val	ttc acg complete Thr of 1950 aaa cca complete Pro of 1965 tca gtc complete Val according to the proof of th	gag gct 7657 Slu Ala gaa tac 7705 Slu Tyr gcg cac 7753
gcg ggg acc can all all at act aga tan act	aa gag gac In Glu Asp 940 ac tct gcc yr Ser Ala tg ata aca eu Ile Thr	gag gcg agc Glu Ala Ser 1949 ccc cct ggg Pro Pro Gly 1960 tca tgc tcc Ser Cys Ser 1975 gtg tac tat Val Tyr Tyr	cta cgg gcc Leu Arg Ala gac ccg ccc Asp Pro Pro tcc aat gtg Ser Asn Val 1980	ttc acg c Phe Thr 6 1950 aaa cca c Lys Pro 6 1965 tca gtc c Ser Val A	gag gct 7657 Slu Ala gaa tac 7705 Slu Tyr gcg cac 7753 Ala His
gcg ggg acc can all all all act aga tag aga tag aga tag aga tag aga tag act aga acc ag	aa gag gac In Glu Asp 940 ac tct gcc yr Ser Ala tg ata aca eu Ile Thr gc aaa agg ly Lys Arg 1990	gag gcg agc Glu Ala Ser 1949 ccc cct ggg Pro Pro Gly 1960 tca tgc tcc Ser Cys Ser 1975 gtg tac tat Val Tyr Tyr	cta cgg gcc Leu Arg Ala gac ccg ccc Asp Pro Pro tcc aat gtg Ser Asn Val 1980 ctc acc cgt Leu Thr Arg 1995 gct aga cac	ttc acg con the Thr Pro V	gag gct 7657 Slu Ala gaa tac 7705 Slu Tyr Gcg cac 7753 Ala His acc acc 7801 Thr Thr 2000 gtc aat 7849

atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa 7945 Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln 2035 2040 2045	5
ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att 7993 Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile- 2050 2055 2060	3
gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt agc 8043 Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser 2065 2070 2075 . 2080	1
gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct 8089 Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala 2085 2090 2095	9
tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat 8137 Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His 2100 2105 2110	7
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct 8185 Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala 2115 2120 2125	5
gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc 8233 Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu 2130 2135 2140	3
aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg 8281 Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp 2145 2150 2155 2160	L
ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt 8329 Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg 2165 2170 2175	€
gcc cga ccc cgc tgg ttc acg tgg tgc cta ctc cta ctt tct gta ggg 8377 Ala Arg Pro Arg Trp Phe Thr Trp Cys Leu Leu Leu Leu Ser Val Gly 2180 2185 2190	7
gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca 8427 Val Gly Ile Tyr Leu Leu Pro Asn Arg * 2195 2200	7
ggccaatagg ccatcetgtt tttttecett ttttecett ttttttttt tttttttt	7 7
<210> 6 <211> 8638 <212> DNA <213> HCV <220> <221> CDS <222> (1802)(8407)	
<400> 6	

		agcco	CC (1acc(JOIOTO	ic: Ota	- 217 / 1						-arai	-ra-	一付つつ(OT DOTA	5 ()	
	T-C1T-T	- ~ ~ ~ ~	100 /										_		_			
																caggac		
																tgccag gccccc		
																gatagg		
																ctaaac		
																gttctc		
																gctgct		
																agaccg		•
																tggcca		
																actggc		
																ccgaga		
•																cctgcc		
																ccggtc		
																tgttcg		
																atgcct		
																gccggc		
																aagagc		
																attcgc		
																tgttaa		
	caga	accac	caa d	eggtt	tccc	et e	tagc	gggat	t caa	attc	cgcc	ccc	cccc	cta a	acgt	tactgg	1260	
	ccga	agco	egc t	tgga	ataa	ag go	ccggt	tgtg	gt1	ttgt	ctat	atgt	tati	ctt (ccac	catatt	1320	
	gccg	jtctt	itt g	gcaa	atgto	ga gg	ggcc	cggaa	a aco	ctgg	ccct	gtct	tctt	tga (cgag	cattcc	1380	
	tagg	ggtc	ett t	cccc	ctctc	g c	caaag	ggaat	gca	aaggi	tctg	ttga	atgi	ccg	tgaag	ggaagc	1440	
																gcagcg		
																tacacc		
																agtcaa		
																ccattg		
																gttaaa		
	laty	gygar	:	•									,					
	aaac	gtct	ag g	geec	ccga	a co	cacgo	ggad	gtg	ggtti	ttcc	tttg	yaaaa	aac a	acgat	taatac	1800	
	aaac c at	gtct g ga	ag g	ia as iccco	ccga ig at	g go	cacgo ca go	gggad ca to	e gtg	ggtti gc gg	ttcc ga gg	tttg gc gc	gaaaa 2g gt	aac a	acgat tc gi	taatac ta ggt	1800 1849	
	aaac c at	gtct g ga et As	ag g	ia as iccco	ccga ig at	a co g go et Al	cacgo ca go	gggad ca to	e gtg	ggtti gc gg ys Gl	ttcc ga gg	tttg gc gc	gaaaa 2g gt	aac a	acgai tc gi he Va	taatac	1800 1849	
	aaac c at Me 1	gtct g ga et As	ag g sp Ai	d eg da eg	eccga ng at nu Me 5	a co g go et Al	cacgg ca gc la Al	gggad ca to la Se	eg to	ggtti gc gg ys Gi	ttcc ga gg ly G: 10	tttg gc gc ly Al	gaaaa eg gt la Va	aac a st't' al P	acgat tc gt he Va	taatac ta ggt al Gly 15	1800 1849	
	aaad c at Me 1 ctg	gtct g ga et As	ag g ac cg sp Ai	geeco gg ga cg G]	eccga ag at .u Me	a co g go et Al	cacgg ca go la Al	gggad ca to la Se	eg to	ggtti gc gg ys Gi :	ttcc ga gg ly G: 10 aag	tttg gc gc ly Al	gaaaa eg gt la Va	ac active al Pl	acgat tc gt he Va	taatac ta ggt al Gly 15 agg	1800 1849	
	aaad c at Me 1 ctg	gtct g ga et As	ag g ac cg sp Ai	geeco gg ga cg G]	eccga ag at .u Me	a co g go et Al	cacgg ca go la Al	gggad ca to la Se	eg to	ggtti gc gg ys Gi :	ttcc ga gg ly G: 10 aag	tttg gc gc ly Al	gaaaa eg gt la Va	ac active al Pl	acgat tc gt he Va	taatac ta ggt al Gly 15	1800 1849	
	aaad c at Me 1 ctg	gtct g ga et As	ag g ac cg sp Ai	geeco gg ga cg G]	eccga ag at .u Me	a co g go et Al	cacgg ca go la Al	gggad ca to la Se	eg to	ggtti gc gg ys Gi :	ttcc ga gg ly G: 10 aag	tttg gc gc ly Al	gaaaa eg gt la Va	ac active al Pl	acgai tc gi he Va gct Ala	taatac ta ggt al Gly 15 agg	1800 1849	
	aaad c at Me 1 ctg	gtct g ga et As	ag g ac cg sp Ai	geeco gg ga gg G] ttg Leu	eccga ag at .u Me	a co g go et Al	cacgg ca go la Al	gggad ca to la Se	eg to	ggtti gc gg ys Gi :	ttcc ga gg ly G: 10 aag	tttg gc gc ly Al	gaaaa eg gt la Va	ac active ctc	acgai tc gi he Va gct Ala	taatac ta ggt al Gly 15 agg	1800 1849	
•	aaac c at Me 1 ctg Leu	gtct g ga t As ata Ile	ag g sp Ai ctc Leu	ttg Leu 20	acc Thr	ttg Leu	tca Ser	gggad ca to la Se ccg Pro	cac His	ggtti gc gg ys G tat Tyr	ttcc ga gg ly G: 10 aag Lys	tttggc gc ly A] ctg Leu	gaaaa eg gt la Va ttc Phe	ac actrical Plant Ctc Leu 30	acgai tc gi he Va gct Ala	taatac ta ggt al Gly 15 agg	1800 1849	
•	aaad c at Me 1 ctg Leu	gtctg gatt As ata Ile	ctc Leu	ttg Leu 20	acc Thr	ttg Leu	tat	gggad ca to la Se ccg Pro	cac His 25	ggtti gc gg ys G tat Tyr	agg	tttggc gc ly A] ctg Leu	gaaaa eg gt la Va ttc Phe	ctc Leu 30	acgai tc gi he Va gct Ala cac	taatac ta ggt al Gly 15 agg Arg	1800 1849 1897	
•	aaad c at Me 1 ctg Leu	gtctg gatt As ata Ile	ctc Leu	ttg Leu 20	acc Thr	ttg Leu	tat	gggad ca to la Se ccg Pro	cac His 25	ggtti gc gg ys G tat Tyr	agg	tttggc gcgly Al	gaaaa eg gt la Va ttc Phe	ctc Leu 30	acgai tc gi he Va gct Ala cac	taatac ta ggt al Gly 15 agg Arg	1800 1849 1897	
	aaad c at Me 1 ctg Leu	gtctg gatt As ata Ile	tgg	ttg Leu 20	acc Thr	ttg Leu	tat	gggad ca to la Se ccg Pro ttt Phe	cac His 25	ggtti gc gg ys G tat Tyr	agg	tttggc gcgly Al	gaaaa g gt la Va ttc Phe gag Glu	ctc Leu 30	acgai tc gi he Va gct Ala cac	taatac ta ggt al Gly 15 agg Arg	1800 1849 1897	
•	aaac c at Me 1 ctg Leu ctc Leu	gtct g ga t As ata Ile	tgg Trp 35	ttg Leu 20	acc Thr	ttg Leu	tat Tyr	ca to la Se ccg Pro ttt Phe 40	cac His 25	ggtti gc gg ys G tat Tyr	agg Arg	tttg gc gc ly Al ctg Leu gcc Ala	ttc Phe gag Glu 45	ctc Leu 30	acgai tc gi he Va gct Ala cac His	taatac ta ggt al Gly 15 agg Arg ttg Leu	1800 1849 1897	
	aaac c at Me 1 ctg Leu ctc Leu	gtct g ga t As ata Ile gtg	tgg tgg	ttg Leu 20 tgg Trp	acc Thr	ttg Leu caa Gln	tca Ser tat Tyr	ca to la Se ccg Pro ttt Phe 40	cac His 25 atc Ile	tat Tyr acc Thr	agg Arg	tttg gc gc ly Al ctg Leu gcc Ala	gaaaa g gt la Va ttc Phe gag Glu 45	ctc Leu 30 gca Ala	acgai tc gi he Va gct Ala cac His	taatac ta ggt al Gly 15 agg Arg ttg Leu	1800 1849 1897	
	aaac c at Me 1 ctg Leu ctc Leu	gtct g ga t As ata Ile gtg Val	tgg tgg	ttg Leu 20 tgg Trp	acc Thr	ttg Leu caa Gln	tat Tyr	ca to la Se ccg Pro ttt Phe 40	cac His 25 atc Ile	tat Tyr acc Thr	agg Arg	tttg gc gc ly Al ctg Leu gcc Ala	gaaaa g gt la Va ttc Phe gag Glu 45	ctc Leu 30 gca Ala	acgai tc gi he Va gct Ala cac His	taatac ta ggt al Gly 15 agg Arg ttg Leu	1800 1849 1897	
	aaac c at Me 1 ctg Leu ctc Leu	gtct g ga t As ata Ile gtg	tgg tgg	ttg Leu 20 tgg Trp	acc Thr	ttg Leu caa Gln	tca Ser tat Tyr	ca to la Se ccg Pro ttt Phe 40	cac His 25 atc Ile	tat Tyr acc Thr	agg Arg	tttg gc gc ly Al ctg Leu gcc Ala	gaaaa g gt la Va ttc Phe gag Glu 45	ctc Leu 30 gca Ala	acgai tc gi he Va gct Ala cac His	taatac ta ggt al Gly 15 agg Arg ttg Leu	1800 1849 1897	
	aaac c at Me 1 ctg Leu ctc Leu	gtct g ga t As ata Ile gtg Val 50	tgg tgg Trp	ttg Leu 20 tgg Trp	acc Thr tta Leu	ttg Leu caa Gln	tat Tyr ctc Leu 55	ca to la Se la Se Pro ttt Phe 40 aac Asn	cac His 25 atc Ile gtt Val	tat Tyr acc Thr	agg Arg	tttg gc gc ly Al ctg Leu gcc Ala ggc Gly 60	ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala	gct Ala gcc Ala	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val	1800 1849 1897 1945	
	aaac c at Me 1 ctg Leu ctc Leu caa Gln	gtct g ga t As ata Ile gtg Val 50	tgg Trp ctc	ttg Leu 20 tgg Trp	acc Thr tta Leu	ttg Leu caa Gln ccc Pro	tat tat Tyr ctc Leu 55	ca to la Se	cac His 25 atc Ile gtt Val	tat Tyr acc Thr	agg Lys agg Arg Gly	tttg gc gc ly Al ctg Leu gcc Ala ggc Gly 60	ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp	gct Ala gcc Ala	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val	1800 1849 1897	
	aaac c at Me 1 ctg Leu ctc Leu caa Gln	gtct g ga t As ata Ile gtg Val 50	tgg Trp ctc	ttg Leu 20 tgg Trp	acc Thr tta Leu	ttg Leu caa Gln ccc Pro	tat tat Tyr ctc Leu 55	ca to la Se	cac His 25 atc Ile gtt Val	tat Tyr acc Thr	agg Arg Gly cta	tttg gc gc ly Al ctg Leu gcc Ala ggc Gly 60	ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp	gct Ala gcc Ala	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr	1800 1849 1897 1945	
	aaac c at Me 1 ctg Leu ctc Leu caa Gln	gtct g ga t As ata Ile gtg Val 50	tgg Trp ctc	ttg Leu 20 tgg Trp	acc Thr tta Leu	ttg Leu caa Gln ccc Pro	tat tat Tyr ctc Leu 55	ca to la Se	cac His 25 atc Ile gtt Val	tat Tyr acc Thr	agg Lys agg Arg Gly	tttg gc gc ly Al ctg Leu gcc Ala ggc Gly 60	ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp	gct Ala gcc Ala	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val	1800 1849 1897 1945	
	ctg Leu cta Gln atc Ile 65	gtct g ga t As ata Ile gtg Val 50 ctc Leu	tgg Trp ctc Leu	ttg Leu 20 tgg Trp atc Ile	acc Thr tta Leu ccc Pro	ttg Leu caa Gln ccc Pro	tat Tyr ctc Leu 55 atc Ile	ca to la Se	cac His 25 atc Ile gtt Val	tat Tyr acc Thr cgg Arg	agg Arg Gly cta Leu 75	tttg gc gc ly Al ctg Leu gcc Ala ggc Gly 60 atc	ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp	gct Ala gcc Ala atc	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80	1800 1849 1897 1945	
	aaac c at Me Leu ctc Leu caa Gln atc Ile 65	gtct g ga t As ata ata Ile gtg Val 50 ctc Leu	tgg Trp ctc Leu	ttg Leu 20 tgg Trp atc Ile acg Thr	acc Thr tta Leu ccc Pro tgc	ttg Leu caa Gln ccc Pro	tca ser tat Tyr ctc Leu 55 atc Ile	ca to la Se	cac His 25 atc Ile gtt Val cca Pro	tat Tyr acc Thr cgg Arg	agg Arg ggg Gly cta Leu 75 atg	tttg gc gg Leu gcc Ala ggc Gly 60 atc Ile	ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp acc Thr	gct Ala atc Ile	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1800 1849 1897 1945	
	ctg Leu cta Gln atc Ile 65	gtct g ga t As ata ata Ile gtg Val 50 ctc Leu	tgg Trp ctc Leu	ttg Leu 20 tgg Trp atc Ile acg Thr	acc Thr tta Leu ccc Pro tgc	ttg Leu caa Gln ccc Pro	tca ser tat Tyr ctc Leu 55 atc Ile	ca to la Se	cac His 25 atc Ile gtt Val cca Pro	tat Tyr acc Thr cgg Arg	agg Arg ggg Gly cta Leu 75 atg	tttg gc gg Leu gcc Ala ggc Gly 60 atc Ile	ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp acc Thr	gct Ala atc Ile	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1800 1849 1897 1945 1993	
	aaac c at Me Leu ctc Leu caa Gln atc Ile 65	gtct g ga t As ata ata Ile gtg Val 50 ctc Leu	tgg Trp ctc Leu	ttg Leu 20 tgg Trp atc Ile acg Thr	acc Thr tta Leu ccc Pro tgc	ttg Leu caa Gln ccc Pro	tca ser tat Tyr ctc Leu 55 atc Ile	ca to la Se	cac His 25 atc Ile gtt Val cca Pro	tat Tyr acc Thr cgg Arg	agg Arg ggg Gly cta Leu 75 atg	tttg gc gg Leu gcc Ala ggc Gly 60 atc Ile	ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp acc Thr	gct Ala atc Ile	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1800 1849 1897 1945 1993	
	aaac c at Me Leu ctc Leu caa Gln atc Ile 65	gtct g ga t As ata ata Ile gtg Val 50 ctc Leu	tgg Trp ctc Leu	ttg Leu 20 tgg Trp atc Ile acg Thr	acc Thr tta Leu Ccc Pro	ttg Leu caa Gln ccc Pro	tca ser tat Tyr ctc Leu 55 atc Ile	ca to la Se	cac His 25 atc Ile gtt Val cca Pro	tat Tyr acc Thr cgg Arg	agg Arg ggg Gly cta Leu 75 atg	tttg gc gg Leu gcc Ala ggc Gly 60 atc Ile	ttc Phe gag Glu 45 cgc Arg	ctc Leu 30 gca Ala gat Asp acc Thr	gct Ala atc Ile gct Ala	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt	1800 1849 1897 1945 1993	
	aaac c at Me 1 ctg Leu caa Gln atc Ile 65 aaa Lys	gtct g ga t As at a at a Ile gtg Val otc Leu at a Ile	tgg Trp ctc Leu	ttg Leu 20 tgg Trp atc Ile acg Thr	acc Thr tta Leu Ccc Pro tgc Cys	ttg Leu caa Gln ccc Pro gcg Ala 70 ata Ile	tat Tyr ctc Leu ctc Leu ctc Leu	ca to la Se	cac His atc Ile gtt Val cca Pro	tat Tyr acc Thr cgg Arg gag Glu ctc Leu 90	agg Gly cta Leu 75 atg Met	tttg gc gg ly Al ctg Leu gcc Ala ggc Gly 60 atc Ile gtg Val	ttc Phe gag Glu 45 cgc Arg ttt Phe ctc Leu	ctc Leu 30 gca Ala gat Asp acc Thr	gct Ala stc Ile gct Ala 95	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly	1800 1849 1897 1945 1993 2041 2089	
	aaac c at Me 1 ctg Leu caa Gln atc Ile 65 aaa Lys	gtct g ga t As at a at a Ile gtg Val 50 ctc Leu atc atc	ctc Leu tgg Trp ctc Leu tag	ttg Leu 20 tgg Trp atc Ile acg Thr ctc Leu gtg	acc Thr tta Leu ccc Pro tgc Cys cca Ala 85 ccg	ttg Leu caa Gln ccc Pro gcg Ala 70 ata Ile	tat Tyr ctc Leu ctc Leu ttc	ca to la Se cog Pro ttt Phe 40 aac Asn cac His	cac His atc Ile gtt Val cca Pro	tat Tyr acc Thr cgg Arg gag Glu ctc Leu 90	agg Gly cta Leu 75 atg Met cac	tttg gc gg ly Al ctg Leu gcc Ala ggc 60 atc Ile gtg Val	ttc Phe gag Glu 45 cgc Arg ttt Phe ctc Leu	ctc Leu 30 gca Ala gat Asp acc Thr cag Gln att	gct Ala atc Ile gct Ala 95 cgt	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly gca	1800 1849 1897 1945 1993	
	aaac c at Me 1 ctg Leu caa Gln atc Ile 65 aaa Lys	gtct g ga t As at a at a Ile gtg Val 50 ctc Leu atc atc	ctc Leu tgg Trp ctc Leu tag	ttg Leu 20 tgg Trp atc Ile acg Thr ctc Leu gtg Val	acc Thr tta Leu ccc Pro tgc Cys cca Ala 85 ccg	ttg Leu caa Gln ccc Pro gcg Ala 70 ata Ile	tat Tyr ctc Leu ctc Leu ttc	ca to la Se cog Pro ttt Phe 40 aac Asn cac His	cac His ate Stale Cyal Ccac Pro	tat Tyr acc Thr cgg Arg gag Glu ctc Leu 90	agg Gly cta Leu 75 atg Met cac	tttg gc gg ly Al ctg Leu gcc Ala ggc 60 atc Ile gtg Val	ttc Phe gag Glu 45 cgc Arg ttt Phe ctc Leu	ctc Leu 30 gca Ala gat Asp acc Thr cag Gln att Ile	gct Ala atc Ile gct Ala 95 cgt Arg	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly gca	1800 1849 1897 1945 1993 2041 2089	
	aaac c at Me 1 ctg Leu caa Gln atc Ile 65 aaa Lys	gtct g ga t As at a at a Ile gtg Val 50 ctc Leu atc atc	ctc Leu tgg Trp ctc Leu tag	ttg Leu 20 tgg Trp atc Ile acg Thr ctc Leu gtg	acc Thr tta Leu ccc Pro tgc Cys cca Ala 85 ccg	ttg Leu caa Gln ccc Pro gcg Ala 70 ata Ile	tat Tyr ctc Leu ctc Leu ttc	ca to la Se cog Pro ttt Phe 40 aac Asn cac His	cac His atc Ile gtt Val cca Pro	tat Tyr acc Thr cgg Arg gag Glu ctc Leu 90	agg Gly cta Leu 75 atg Met cac	tttg gc gg ly Al ctg Leu gcc Ala ggc 60 atc Ile gtg Val	ttc Phe gag Glu 45 cgc Arg ttt Phe ctc Leu	ctc Leu 30 gca Ala gat Asp acc Thr cag Gln att	gct Ala atc Ile gct Ala 95 cgt Arg	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly gca	1800 1849 1897 1945 1993 2041 2089	
	aaac c at Me 1 ctg Leu caa Gln atc Ile 65 aaa Lys	gtct g ga t As at a at a Ile gtg Val 50 ctc Leu atc atc	ctc Leu tgg Trp ctc Leu tag	ttg Leu 20 tgg Trp atc Ile acg Thr ctc Leu gtg Val	acc Thr tta Leu ccc Pro tgc Cys cca Ala 85 ccg	ttg Leu caa Gln ccc Pro gcg Ala 70 ata Ile	tat Tyr ctc Leu ctc Leu ttc	ca to la Se cog Pro ttt Phe 40 aac Asn cac His	cac His ate Stale Cyal Ccac Pro	tat Tyr acc Thr cgg Arg gag Glu ctc Leu 90	agg Gly cta Leu 75 atg Met cac	tttg gc gg ly Al ctg Leu gcc Ala ggc 60 atc Ile gtg Val	ttc Phe gag Glu 45 cgc Arg ttt Phe ctc Leu	ctc Leu 30 gca Ala gat Asp acc Thr cag Gln att Ile	gct Ala atc Ile gct Ala 95 cgt Arg	taatac ta ggt al Gly 15 agg Arg ttg Leu gtc Val acc Thr 80 ggt Gly gca	1800 1849 1897 1945 1993 2041 2089	

	atg Met		_							_	_			•	2185
	atg Met 130			_	_	-		=				_			2233
_	cca Pro					_									2281
	gtt Val														2329
	GJÀ aaa														2377
_	tcc Ser	_			-			_	_	•	 _	_	_		2425
	gaa Glu 210			_							_	_			2473
_	cag Gln	_		_				_							2521
	gac Asp			_					_		 _				2569
	caa Gln														2617
	cat His														2665
	caa Gln 290														2713
	ccc Pro											_	_		2761
	tac Tyr						_		_			_			2809

						cta Leu									ttg Leu	2857
						cca Pro									gtg Val	2905
Gly	atc Ile 370	Phe	cgg	gct Ala	gcc Ala	gtg Val 375	tgc Cys	acc Thr	cga Arg	Gly 999	gtt Val 380	gcg Ala	aag Lys	gcg Ala	gtg Val	2953
gac Asp 385	ttt Phe	gta Val	ccc Pro	gtc Val	gag Glu 390	tct Ser	atg Met	gga Gly	acc Thr	act Thr 395	atg Met	cgg Arg	tcc Ser	ccg Pro	gtc Val 400	3001
						cct Pro									gtg Val	3049
Ala	His	Leu	His 420	Ala	Pro	Thr	Gly	Ser 425	Gly	Lys	Ser	Thr	Lys 430	Val		3097
Ala	Ala	Tyr 435	Ala	Ala	Gln	Gly	Tyr 440	Lys	Val	Leu	Val	Leu 445	Asn	Pro		3145
Val	Ala 450	Ala	Thr	Leu	Gly	Phe 455	Gly	Ala	Tyr	Met	Ser 460	Lys	Ala	His	_	3193
Ile 465	Asp	Pro	Asn	Ile	Arg 470	acc Thr	Gly	Val	Arg	Thr 475	Ile	Thr	Thr	Gly	Ala 480	3241
Pro	Ile	Thr	Tyr	Ser 485	Thr	tat Tyr	Gly	Lys	Phe 490	Leu	Ala	Asp	Gly	Gly 495	Cys	3289
Ser	Gly	Gly	Ala 500	Tyr	Asp	atc Ile	Ile	Ile 505	Cys	Asp	Glu	Cys	His 510	Ser	Thr	3337
gac Asp	tcg Ser	acc Thr 515	act Thr	atc Ile	ctg Leu	ggc	atc Ile 520	ggc	aca Thr	gtc Val	ctg Leu	gac Asp 525	caa Gln	gcg Ala	gag Glu	3385
acg Thr	gct Ala 530	gga Gly	gcg Ala	cga Arg	ctc Leu	gtc Val 535	gtg Val	ctc Leu	gcc Ala	acc Thr	gct Ala 540	acg Thr	cct Pro	ccg Pro	gga Gly	3433
tcg Ser 545	gtc Val	acc Thr	gtg Val	cca Pro	cat His 550	cca Pro	aac Asn	atc Ile	gag Glu	gag Glu 555	gtg Val	gct Ala	ctg Leu	tcc Ser	agc Ser 560	3481

	gga Gly	_		•					_			_			3529
-	gjå aaà			_						_				gat Asp	3577
_	ctc Leu		_			_					_		_		3625
, tac Tyr	cgg Arg 610			_	_	_				_		_	_		3673
- -	gta Val	_				_	_					_		gac Asp 640	3721
	gtg Val			- - -		_	_		_		_			agc Ser	3769
	gac Asp						-								3817
_	tca Ser			_		_		_		· ·					3865
	tac Tyr 690													gat Asp	3913
	tcg Ser							_						gag Glu 720	3961
	acg Thr							Leu		Ala	Tyr	Leu			4009
	gly aaa							_			_			_	4057
	aca Thr						•			_		_		_	4105
	gca Ala 770														4153

												atg Met		_	4201
	-					_		_				ccc Pro	_	ctg Leu	4249
		_	_	_	_							cac His 830			4297
												gtc Val			4345
											_	gcc Ala			4393
												atc Ile	Leu		4441
												cgg Arg		ttc Phe	4489
_												gaa Glu 910		gga Gly	4537
												ttg Leu		caa Gln	4585
												gaa Glu		aag Lys	4633
												aat Asn		atc Ile 960	4681
							- -			_		Gly			4729
	_		_	_				_				agc Ser 990	_		4777
		His			_		Asn		_			tgg Trp		_	4825

	c caa a Gln 1010	Leu					Ala					Val				4873
Il	c gct e Ala 25					Gly					Gly					4921
	t att p Ile				Tyr			_	_	Ala	_			_	Ala	4969
	t aag e Lys	_		Ser					Ser					Val		5017
	a ctc u Leu		Ala				•	Gly	_			-	Gly	=	=	5065
	c gca s Ala 1090	Ala		_	_		His	-			T.	Glu		_	•	5113
Gl	g tgg n Trp 05	_			_	Ile			—		Arg			_	_	5161
	c ccc r Pro				Val		_	_	_	Ala	_		_		Thr	5209
	g atc n Ile			Ser					Gln	_	_	-		Leu		5257
	g tgg n Trp		Asn		_	_		Thr				- -	Ser			5305
	a gat g Asp 1170	Val		_		Ile	Cys		Val	Leu	_	Asp			_	5353
Tr	g ctc p Leu 85	_				Leu			=	_	Gly	_		_		5401
	a tgt r Cys	_		_	Tyr	_	_ -	_		Arg	_		_	_	Met	5449
ca	2 200	200	taa	999	tat	aas	gca	cad	atc	200	aas	cat	ata	aaa	220	5497

		agg acc tgt agt a Arg Thr Cys Ser A 1	· -
		acc acg ggc ccc t Thr Thr Gly Pro C 1260	
		ctg tgg cgg gtg g Leu Trp Arg Val A 1275	- -
		gat ttc cac tac g Asp Phe His Tyr V 1290	
	Val Lys Cys Pro	tgt cag gtt ccg g Cys Gln Val Pro A 1305	_
		ttg cac agg tac g Leu His Arg Tyr A 1	
		aca ttc ctg gtc g Thr Phe Leu Val G 1340	-
		tgc gag ccc gaa c Cys Glu Pro Glu P 1355	
		ccc tcc cac att a Pro Ser His Ile T 1370	
_	Leu Ala Arg Gly	tct ccc ccc tcc t Ser Pro Pro Ser L 1385	
Ser Ala Ser Gln		tcc ttg aag gca a Ser Leu Lys Ala T 1	_
_		ctc atc gag gcc a Leu Ile Glu Ala A 1420	
		acc cgc gtg gag t Thr Arg Val Glu S 1435	
Val Val Ile Leu	gac tct ttc gag Asp Ser Phe Glu 1445	ccg ctc caa gcg g Pro Leu Gln Ala G 1450	ag gag gat gag 6169 lu Glu Asp Glu 1455

	-			Val					ctg Leu					Lys	_	6217
			Met		•			Arg	ccg Pro				Pro			6265
	-	Ser					Asp		gtc Val			Val				6313
_	Pro				_	Lys	_		ccg Pro		Pro					6361
=					Leu	•			acc Thr 1530	Val					Ala	6409
-		=		Lys					tcc Ser		-		_	Val		6457
_		_	Ala					Asp	cag Gln			· — ·	Asp		_	6505
		Ser		_		_	Tyr		tcc Ser	_		Pro				6553
	Pro		-		-	Leu			gjå aaa		Trp			_	_	6601
	- -	-	-		Asp			-	tgc Cys 1610	Ser	_			_	Trp	6649
		_		Ile	Thr	Pro	Cys	Ala	gcg Ala	Glu	Glu	Thr	Lys	Leu		6697
			Leu					Leu	cgt Arg				Leu	_		6745
		Thr		_		_	Ser	_	cgg Arg	_	_	Lys		_	_	6793
	Arg			_	_	Asp	_	_	tac Tyr		Asp			_		6841

	aag Lys				Ser					Lys				_		6889
	gcc Ala			Leu					Ser					Phe	ggc Gly	6937
	GJA 333	_	Lys		_			Leu				_	Val		cac His	6985
_	cgc Arg 1730	Ser					Leu					Glu			att Ile	7033
	Thr					ГЛЗ		-			Cys	_			gag Glu 1760	. 7081
	gjå aaa				Pro					Val				_	_	7129
_	cgt Arg	_		Glu					Tyr	•				Thr	ctc Leu	7177
	cag Gln		Val					Tyr					Ser			7225
	cgg Arg 1810	Val					Asn					Lys				7273
	Gly					Thr					Ser				gag Glu 1840	7321
	gac Asp				Glu					Gln		Cys	Asp		Ala	7369
															400	7417
	gaa Glu			Gln					Leu					Tyr		741
		Ala	Arg 1860 ctg Leu	Gln) act	Ala aat	Ile tct	Arg aaa	Ser 1869 ggg Gly	Leu cag	Thr	Glu	Arg ggc	Leu 1870 tat Tyr	Tyr) cgc	Ile cgg	7465

tgt tac t Cys Tyr L 1905		Ala A			_			Ala				_	7561
tgc acg a Cys Thr M		_					Val	_			_	Ser	7609
gcg ggg a	_	Glu A				Leu			_	_	Glu		7657
atg act ag Met Thr A		_			Gly					Pro	_		7705
gac ttg g Asp Leu G 1970				Cys					Ser				7753
gat gca to Asp Ala So 1985		Lys A						Arg					7801
ccc ctt g		-				_	Arg				_	Asn	7849
tcc tgg c Ser Trp L	· —	Asn I				Ala			_		Ala		7897
atg atc c Met Ile L 2					Ser					Gln			7945
ctt gaa a Leu Glu L 2050	_	_	_	Gln				_	Cys				7993
gag cca c Glu Pro L 2065		Leu P	_	Ile	Ile	Gln	Arg	Leu	His	Gly	Leu		8041
gca ttt to Ala Phe So							Glu			· —		Ala	8089
tca tgc co		Lys L		_	•	Pro	_		_		Arg	_	8137
cgg gcc ag Arg Ala A		_			Leu					Gly			8185

gcc act tgt Ala Thr Cys 2130	- -	-		Phe			_		Arg				8233
aaa ctc act Lys Leu Thr 2145		Ile Pı				=	_	Asp			-		8281
ttc gtt gct Phe Val Ala	Gly S				-		Tyr	_	_	 -		Arg	8329
gcc cga ccc Ala Arg Pro				Trp	=	Leu					Val		8377
gta ggc atc Val Gly Ile 219	Tyr 1				Arg		acgg	gggag	gct a	aaac	actco	ca ·	8427
ggccaatagg		_											
ttttttttcccatcttagc	_											-	
gtgctgatac						-55	, , ,	J~J`					8638
<210> 7 <211> 8638 <212> DNA <213> HCV <220>													
<221> CDS													
<222> (1802) { {	8407)	•							•			
<400> 7							•						
gccagcccc			_										
tcttcacgca cccccctccc												-	
gacgaccggg			_		_								
gcgagactgc											_		
gtgcttgcga													
ctcaaagaaa	aaccaa	aaggg	cgcgc	catga	tte	gaaca	aga	tgga	attgo	cac	gcag	gttctc	420
cggccgcttg									-			_	
ctgatgccgc													
acctgtccgg cgacgggcgt													
tgctattggg													
aagtatccat													
cattcgacca				_									
ttgtcgatca												_	
ccaggeteaa													
gcttgccgaa													
tgggtgtggc ttggcggcga													
agcgcatcgc													
cagaccacaa													
ccgaagccgc				_	_	=							
	_					_							
gccgtctttt	ggcaai	tgtga	gggcc	cggaa	aco	ctggo	ccct	gtc	ttcti	tga	cgag	cattcc	1380

taggggtctt taggttcctctg gaaccccca caggaaaggcg gatggctctcc tatgggatct gaaaggtctag gaaaaggtctag gaaaggtctag gaaaggtctag gaaaaggtctag gaaaaggtcag gaaaaggtctag gaaaaggtctag gaaaaggtctag gaaaaggtctag gaaaaggtag gaaaaggtctag gaaaaggtctag gaaaaggtctag gaaaaggtctag gaaaaggtctag gaaaaggtcag gaaaa	aagettett ga etggegaea ga eacaacee aa eaagegtat to atetgggge ca ecceegaa e	agacaaac agtgcctctg cgtgccacgt tgaacaaggg gtcacaaggg gtcacaaggg gtcacaagggac gcacaggggac g	acgtctgta ggccaaaag gtgagttgg ctgaaggat tgctttaca tggttttcc tgc gga g	gcgacccttt ccacgtgtat atagttgtgg gcccagaagg tgtgtttagt tttgaaaaac gc gcg gtt t	gcaggcagcg aagatacacc aaagagtcaa taccccattg cgaggttaaa acgataatac tc gta ggt	1500 1560 1620 1680 1740 1800
ctg ata ctc Leu Ile Leu		_	s Tyr Lys	-	Ala Arg	1897
ctc ata tgg Leu Ile Trp 35		_			_	1945
caa gtg tgg Gln Val Trp 50						1993
atc ctc ctc Ile Leu Leu . 65						2041
aaa atc ttg Lys Ile Leu	_		-	- -		2089
ata acc aaa Ile Thr Lys	- - -	-	g Ala His		Arg Ala	2137
tgc atg ctg Cys Met Leu 115			_			2185
ctc atg aag Leu Met Lys 130				_		2233
acc cca ctg Thr Pro Leu 145						2281
gca gtt gag Ala Val Glu		-				2329
tgg ggg gca Trp Gly Ala			y Asp Ile		Leu Pro	2377
gtc tcc gcc Val Ser Ala 195						2425

	gaa Glu 210					_							tcc Ser	2473
_	cag Gln	_						_			_		ggc Gly 240	2521
	gac Asp			_	_	-		_					gca Ala	2569
	caa Gln												gtc Val	2617
	cat His	_	_	_										2665
	caa Gln 290												gcg Ala	2713
	ccc Pro													2761
	tac Tyr													2809
_	gac Asp			_									ttg Leu	2857
	Gly											_	gtg Val	2905
	atc Ile 370													2953
	ttt Phe						_	_		_		_	_	3001
_	acg Thr						=	-				_	_	3049
	cat His													3097

	gcg Ala							Lys							tcc Ser	3145
	gcc Ala 450	_						_					_		ggt Gly	3193
	gac Asp							_						_	gcc Ala 480	3241
	atc Ile				_						_				tgc Cys	3289
	Gly				•				_		_		_			3337
-	tcg Ser															3385
_			_	_		_	_		_		_			_	gga Gly	3433
_	gtc Val												-		_	3481
	gga Gly	_							_							3529
_	el aaa							_				_			_	3577
_	ctc Leu	-		-	-	Ser	Gly	Leu		Leu	Asn	-	_	_		3625
	cgg Arg 610			-			-									3673
	gta Val								-							3721
	gtg Val						_	_								3769

		ccg Pro													gcg Ala	3817
		cgc Arg 675													ggc Gly	3865
att Ile	tac Tyr 690	agg Arg	ttt Phe	gtg Val	act Thr	cca Pro 695	gga Gly	gaa Glu	cgg Arg	ccc Pro	tcg Ser 700	gly	atg Met	ttc Phe	gat Asp	3913
tcc Ser 705	tcg Ser	gtt Val	ctg Leu	tgc Cys	gag Glu 710	tgc Cys	tat Tyr	gac Asp	gcg Ala	ggc Gly 715	tgt Cys	gct Ala	tgg Trp	tac Tyr	gag Glu 720	3961
ctc Leu	acg Thr	ccc Pro	gcc Ala	gag Glu 725	acc Thr	tca Ser	gtt Val	agg Arg	ttg Leu 730	cgg Arg	gct Ala	tac Tyr	cta Leu	aac Asn 735	aca Thr	4009
Pro	Gly	Leu	Pro 740	Val	Cys	Gln	Asp	His 745	Leu	Glu	Phe	Trp	Glu 750	Gly		
Phe	Thr	Gly 755	Leu	Thr	His	Ile	Asp 760	Ala	His	Phe	Leu	Ser 765	Gln	Thr	aag Lys	4105
cag Gln	gca Ala 770	gga Gly	gac Asp	aac Asn	ttc Phe	ccc Pro 775	tac Tyr	ctg Leu	gta Val	gca Ala	tac Tyr 780	cag Gln	gct Ala	acg Thr	gtg Val	4153
tgc Cys 785	'gcc Ala	agg Arg	gct Ala	cag Gln	gct Ala 790	cca Pro	cct Pro	cca Pro	tcg Ser	tgg Trp 795	gac Asp	caa Gln	atg Met	tgg Trp	aag Lys 800	4201
		ata Ile													ctg Leu	4249
		ctg Leu		Ala		Gln	Asn		Val		Thr	Thr				4297
acc Thr	aaa Lys	tac Tyr 835	atc Ile	atg Met	gca Ala	tgc Cys	atg Met 840	tcg Ser	gct Ala	gac Asp	ctg Leu	gag Glu 845	gtc Val	gtc Val	acg Thr	4345
agc Ser	acc Thr 850	tgg Trp	gtg Val	ctg Leu	gta Val	ggc Gly 855	gga Gly	gtc Val	cta Leu	gca Ala	gct Ala 860	ctg Leu	gct Ala	gcg Ala	tat Tyr	4393
tgc Cys 865	ctg Leu	aca Thr	aca Thr	ggc	agc Ser 870	gtg Val	gtc Val	att Ile	gtg Val	ggc Gly 875	agg Arg	atc Ile	atc Ile	ttg Leu	tcc Ser 880	4441

		att ccc g Ile Pro A		Val Leu			4489
		tgt gcc t Cys Ala S			-		4537
Met Gln L		caa ttc a Gln Phe L 9					4585
		gcg gag g Ala Glu A 935			Val Glu		4633
		gcc ttc t Ala Phe T 950					4681
		tta gca g Leu Ala G		Thr Leu			4729
		atg gca t Met Ala P					4777
Thr Thr G		ctc ctg t Leu Leu P					4825
		ccc agc g Pro Ser A 1015			Val Gly		4873
		gtt ggc ag Val Gly S 1030					4921
		tat gga g Tyr Gly A 5		. Ala Gly			4969
		ggc gag a Gly Glu M	_			Val Asn	5017
Leu Leu Pr		ctc tcc control Leu Ser P.		_			5065
tgc gca go Cys Ala Al	cg ata ctg	cgt cgg c				-	5113

cag Gln 1105	Trp	atg Met	aac Asn	cgg Arg	ctg Leu 111(Ile	gcg Ala	ttc Phe	gct Ala	tcg Ser 111!	Arg	ggt Gly	aac Asn	cac His	gtc Val 1120	5161
tcc Ser	ccc Pro	acg Thr	cac His	tat Tyr 1125	Val	cct Pro	gag Glu	agc Ser	gac Asp 1130	Ala	gca Ala	gca Ala	cgt Arg	gtc Val 1135		5209
cag Gln	atc Ile	ctc Leu	tct Ser 1140	Ser	ctt Leu	acc Thr	atc Ile	act Thr 1149	Gln	ctg Leu	ctg Leu	aag Lys	agg Arg 1150	Leu	cac His	5257
cag Gln	tgg Trp	atc Ile 1155	Asn	gag Glu	gac Asp	tgc Cys	tcc Ser 1160	Thr	cca Pro	tgc Cys	tcc Ser	ggc Gly 1169	Ser	tgg Trp	cta Leu	5305
Arg .	Asp 117(Val)	Trp	Asp	Trp	Ile 1175	Cys	Thr	Val	Leu	Thr 1180	Asp)	Phe	Lys		5353
Trp 1	Leu	Gln	Ser	Lys	Leu 1190	Leu	Pro	Arg	Leu	Pro 1195	Gly	Val	Pro	Phe	1200	5401
Ser	Cys	Gln	Arg	Gly 1205	Tyr	Lys	Gly	Val	Trp 1210	Arg)	Gly	Asp	Gly	Ile 1215	5	5449
caa (Gln '	Thr	Thr	Cys 1220	Pro	Cys	Gly	Ala	Gln 1225	Ile	Thr	Gly	His	Val 1230	Lys)	Asn	5497
Cys (Ser	Met 1235	Arg	Ile	Val	Gly	Pro 1240	Arg	Thr	Cys	Ser	Asn 1245	Thr	Trp		5545
Gly :	Thr 1250	Phe	Pro	Ile	Asn	Ala 1255	Tyr	Thr	Thr	Gly	Pro 1260	Cys)	Thr	Pro		5593
Pro 1	Ala	Pro	Asn	Tyr	Ser 1270	Arg	Ala	Leu	Trp	Arg 1275	Val	Ala	Ala	Glu	Glu 1280	5641
tac (gtg Val	gag Glu	Val	acg Thr 1285	Arg	gtg Val	gly aaa	Asp	ttc Phe 1290	His	tac Tyr	gtg Val	Thr	ggc Gly 1295	Met	5689
acc a	act Thr	Asp .	aac Asn 1300	Val	řì\a ř	tgc Cys	Pro	tgt Cys 1305	Gln	gtt Val	ccg Pro	gcc Ala	ccc Pro 1310	Glu	ttc Phe	5737
ttc a																

Lys Pro Leu Leu 1330		Val Thr Phe			5833
tac ccg gtt ggg Tyr Pro Val Gly 1345		Pro Cys Glu			5881
gtg ctc act tcc Val Leu Thr Ser	_		His Ile Thr	- - -	5929
gct aag cgt agg Ala Lys Arg Arg 138	Leu Ala Arg				5977
tca gct agc cag Ser Ala Ser Glr 1395	Leu Ser Ala			Cys Thr Thr	6025
cgt cat gac tcc Arg His Asp Ser 1410		Asp Leu Ile			6073
cgg cag gag atg Arg Gln Glu Met 1425		Ile Thr Arg			6121
gta gta att tto Val Val Ile Lev	-	- -	Gln Ala Glu		6169
-	Asp Ser Phe 1445 gtt ccg gcg Val Pro Ala	Glu Pro Leu 1450 gag atc ctg	Gln Ala Glu cgg agg tcc	Glu Asp Glu 1455 agg aaa ttc	6169 6217
Val Val Ile Lev agg gaa gta tco Arg Glu Val Ser	Asp Ser Phe 1445 gtt ccg gcg Val Pro Ala 0 ccc ata tgg	Glu Pro Leu 1450 gag atc ctg Glu Ile Leu 1465 gca cgc ccg	cgg agg tcc Arg Arg Ser	Glu Asp Glu 1455 agg aaa ttc Arg Lys Phe 1470 cct cca ctg Pro Pro Leu	
agg gaa gta tcc Arg Glu Val Ser 146 cct cga gcg atc Pro Arg Ala Met 1475 tta gag tcc tgc Leu Glu Ser Try	Asp Ser Phe 1445 gtt ccg gcg Val Pro Ala 0 ccc ata tgg Pro Ile Trp aag gac ccg Lys Asp Pro	Glu Pro Leu 1450 gag atc ctg Glu Ile Leu 1465 gca cgc ccg Ala Arg Pro 1480 gac tac gtc	cgg agg tcc Arg Arg Ser gat tac aac Asp Tyr Asn 1483 cct cca gtg Pro Pro Val	Glu Asp Glu 1455 agg aaa ttc Arg Lys Phe 1470 cct cca ctg Pro Pro Leu 5	6217
agg gaa gta tcc Arg Glu Val Ser 146 cct cga gcg atc Pro Arg Ala Met 1475 tta gag tcc tgg Leu Glu Ser Try	Asp Ser Phe 1445 gtt ccg gcg Val Pro Ala 0 ccc ata tgg Pro Ile Trp aag gac ccg Lys Asp Pro 1495	Glu Pro Leu 1450 gag atc ctg Glu Ile Leu 1465 gca cgc ccg Ala Arg Pro 1480 gac tac gtc Asp Tyr Val gcc cct ccg Ala Pro Pro	cgg agg tcc Arg Arg Ser gat tac aac Asp Tyr Asn 1483 cct cca gtg Pro Pro Val 1500 ata cca cct	Glu Asp Glu 1455 agg aaa ttc Arg Lys Phe 1470 cct cca ctg Pro Pro Leu 5 gta cac ggg Val His Gly cca cgg agg	621 7 6265
agg gaa gta tcc Arg Glu Val Ser 146 cct cga gcg atc Pro Arg Ala Met 1475 tta gag tcc tgg Leu Glu Ser Try 1490 tgt cca ttg ccg Cys Pro Leu Pro	Asp Ser Phe 1445 gtt ccg gcg Val Pro Ala 0 ccc ata tgg Pro Ile Trp aag gac ccg Lys Asp Pro 1495 cct gcc aag Pro Ala Lys 1510 gtc ctg tca	gag atc ctg Glu Ile Leu 1465 gca cgc ccg Ala Arg Pro 1480 gac tac gtc Asp Tyr Val gcc cct ccg Ala Pro Pro	cgg agg tcc Arg Arg Ser gat tac aac Asp Tyr Asn 1483 cct cca gtg Pro Pro Val 1500 ata cca cct Ile Pro Pro 1515 gtg tct tct Val Ser Ser	Glu Asp Glu 1455 agg aaa ttc Arg Lys Phe 1470 cct cca ctg Pro Pro Leu 5 gta cac ggg Val His Gly cca cgg agg Pro Arg Arg 1520 gcc ttg gcg	6217 6265 6313

Ser Gly Thr	gca acg gcc to Ala Thr Ala Se 5	t cct gac ca r Pro Asp Gl 1560	ag ccc tcc gac In Pro Ser Asp 156	Asp Gly Asp
	gac gtt gag to Asp Val Glu Se 15			
gag ccg ggg Glu Pro Gly 1585	gat ccc gat ct Asp Pro Asp Le 1590	c agc gac go u Ser Asp Gl	gg tct tgg tct ly Ser Trp Ser 1595	acc gta agc 6601 Thr Val Ser 1600
gag gag gct Glu Glu Ala	agt gag gac gt Ser Glu Asp Va 1605	l Val Cys Cy	gc tcg atg tcc ys Ser Met Ser 510	tac aca tgg 6649 Tyr Thr Trp 1615
Thr Gly Ala	ctg atc acg cc Leu Ile Thr Pr 1620	o Cys Ala Al 1625	la Glu Glu Thr	Lys Leu Pro 1630
Ile Asn Ala 163		r Leu Leu Ar 1640	rg His His Asn 164!	Leu Val Tyr
gct aca aca Ala Thr Thr 1650	tet ege age ge Ser Arg Ser Al 16	a Ser Leu Ar	gg cag aag aag cg Gln Lys Lys 1660	gtc acc ttt 6793 Val Thr Phe
gac aga ctg Asp Arg Leu 1665	cag gtc ctg ga Gln Val Leu As 1670	c gac cac ta p Asp His Ty	r Arg Asp Val 1675	ctc aag gag 6841 Leu Lys Glu 1680
atg aag gcg Met Lys Ala	aag gcg tcc ac Lys Ala Ser Th 1685	r Val Lys Al	t aaa ctt cta a Lys Leu Leu 590	tcc gtg gag 6889 Ser Val Glu 1695
gaa gcc tgt Glu Ala Cys	aag ctg acg cc Lys Leu Thr Pr 1700	c cca cat to Pro His Se 1705	eg gcc aga tct er Ala Arg Ser	aaa ttt ggc 6937 Lys Phe Gly 1710
tat ggg gca Tyr Gly Ala 171	aag gac gtc cg Lys Asp Val Ar 5	g aac cta to g Asn Leu Se 1720	er Ser Lys Ala	Val Asn His
atc cgc tcc Ile Arg Ser 1730	gtg tgg aag ga Val Trp Lys As 17	p Leu Leu Gl	na gac act gag .u Asp Thr Glu 1740	aca cca att 7033 Thr Pro Ile
gac acc acc Asp Thr Thr 1745	atc atg gca aa Ile Met Ala Ly 1750	a aat gag gt 3 Asn Glu Va	t ttc tgc gtc il Phe Cys Val 1755	caa cca gag 7081 Gln Pro Glu 1760
	cgc aag cca gc Arg Lys Pro Ala 1765	a Arg Leu Il		

gtt cgt gtg (Val Arg Val (Tyr Asp Val		
cct cag gcc g Pro Gln Ala 1 1795	Val Met Gly				
cag cgg gtc g Gln Arg Val (1810	_			Lys Lys Cys	
atg ggc ttc g Met Gly Phe 1 1825		Thr Arg Cys			
aat gac atc o Asn Asp Ile i					Ala .
ccc gaa gcc a Pro Glu Ala			Leu Thr Glu		
ggg ggc ccc (Gly Gly Pro 1 1875	Leu Thr Asn				
tgc cgc gcg a Cys Arg Ala a 1890	_			Asn Thr Leu	_
tgt tac ttg : Cys Tyr Leu : 1905		Ala Ala Cys	_	_	
tgc acg atg (Cys Thr Met)					Ser
gcg ggg acc (Ala Gly Thr			Leu Arg Ala		
atg act aga Met Thr Arg 1 1955	Tyr Ser Ala	_ - -	-	-	
gac ttg gag Asp Leu Glu 1 1970		-	~ ~	Ser Val Ala	
gat gca tct g Asp Ala Ser (1985		Val Tyr Tyr	_	_	

ccc ctt gcg cg		<u> </u>			-	7849
tcc tgg cta gg Ser Trp Leu Gl 20	y Asn Ile Ile	_	Ala Pro	Thr Leu Tr		7897
atg atc ctg at Met Ile Leu Me 2035				-		7945
ctt gaa aaa gc Leu Glu Lys Al 2050		Gln Ile				7993
gag cca ctt ga Glu Pro Leu As 2065	_	e '	_	Leu His G		8041
gca ttt tca ct Ala Phe Ser Le					•	8089
tca tgc ctc ag Ser Cys Leu Ar 21	g Lys Leu Gly	_	Pro Leu	Arg Val Tr	_	8137
cgg gcc aga ag Arg Ala Arg Se 2115				- -	- -	8185
gcc act tgt gg Ala Thr Cys Gl 2130		Phe Asn	_			8233
aaa ctc act cc Lys Leu Thr Pr 2145				Asp Leu Se		8281
ttc gtt gct gg Phe Val Ala Gl					-	8329
gcc cga ccc cg Ala Arg Pro Ar	g Trp Phe Met		Leu Leu	Leu Leu Se		8377
gta ggc atc ta Val Gly Ile Ty 2195.				ggagct aaa	acactcca	8427
ggccaatagg cca ttttttttt ttt ccatcttagc cct gtgctgatac tgg	tctcctt ttttt agtcacg gctag	ttcct ctt	tttttcc	ttttctttcc	: tttggtggct	8547

<211><212><213>	DNA						
<400> accago						•	6
<210><211><211><212><213>	63 DNA		•				
<400> gaatte gac		tggcgcgccc	agatgttaac	cagatccatg	gcacactcta	gagtactgtc	60 63
<210><211><212><213>	33 DNA						
<400> cggaat		aacagaccac	aacggtttcc	ctc			33
<210><211><212><213>	30 DNA		•			-	
<400> ggcgta		tggtattatc	gtgtttttca				30
<210><211><212><212><213>	45 DNA						
<400> gcatat		tctaatacga	ctcactatag	gccagccccc	gattg	•	45
<210><211><211><212><213>	45 DNA						
<400> ggcgcg		ttggtttttc	tttgaggttt	aggattcgtg	ctcat		45
<210><211><211><212><213>	36 DNA						
<400> aaaggg		tgattgaaca	agatggattg	cacgca		•	36
<210> <211>							

<212> DNA <213> HCV	
<400> 15 gcatatgtta actcagaaga actcgtcaag aaggcgata	39
<210> 16 <211> 45 <212> DNA <213> HCV	
<400> 16	
gcatatgaat tctaatacga ctcactatag gccagccccc gattg	45
<210> 17 <211> 30 <212> DNA <213> HCV	
<400> 17	
acgcagaaag cgtctagcca tggcgttagt	30
<pre> <210> 18 <211> 30 <212> DNA <213> HCV </pre>	
<400> 18	
tcccggggca ctcgcaagca ccctatcagg	30
<210> 19 <211> 26 <212> DNA <213> HCV	
<220>	
<223> Label with FAM: fluorescence reporter dye	
<223> Label with TAMRA: Quencher dye	
<400> 19	
tggtctgcgg aacgggtgag tacacc	26
<210> 20 <211> 45 <212> DNA <213> HCV	
<400> 20	
gtggacgaat tctaatacga ctcactataa ccagcccccg attgg	45
<210> 21 <211> 27 <212> DNA <213> HCV	
<400> 21	
ggaacgcccg tcgtggccag ccacgat	27

```
<210> 22
<211> 23
<212> DNA
<213> HCV
<400> 22
gtcgtcttct ctgacatgga gac
                                                                  23
<210> 23
<211> 27
<212> DNA
<213> HCV
<400> 23
gagttgctca gtggattgat gggcagc
                                                                  27
<210> 24
<211> 8638
<212> DNA
<213> HCV
<220>
<221> CDS
<222> (1802)...(8407)
<400> 24
accagecece gattggggge gaeacteeae catagateae teeeetgtga ggaactaetg 60
tetteaegea gaaagegtet agecatggeg ttagtatgag tgtegtgeag ceteeaggae 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg teetttettg gateaacccg etcaatgeet ggagatttgg gegtgeecc 240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtatccat catggctgat gcaatgcggc ggctgcatac gcttgatccg gctacctgcc 780
cattcgacca ccaagcgaaa catcgcatcg agcgagcacg tactcggatg gaagccggtc 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggctcaa ggcgcgcatg cccgacggcg aggatctcgt cgtgacccat ggcgatgcct 960
gcttgccgaa tatcatggtg gaaaatggcc gcttttctgg attcatcgac tgtggccggc 1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc 1140
agegeatege ettetatege ettettgaeg agttettetg agttegegee eagatgttaa 1200
cagaccacaa cggtttccct ctagcgggat caattccgcc cccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
gccgtctttt ggcaatgtga gggcccggaa acctggccct gtcttcttga cgagcattcc 1380
taggggtctt tcccctctcg ccaaaggaat ġcaaggtctg ttgaatgtcg tgaaggaagc 1440
agttcctctg gaagcttctt gaagacaaac aacgtctgta gcgacccttt gcaggcagcg 1500
gaaccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggctctcc tcaagcgtat tcaacaaggg gctgaaggat gcccagaagg taccccattg 1680
tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740
aaacgtctag gcccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
```

	et A	sp A	rg G		et Ai 5	la A	la Se	er Cy		ly G) 10	ly Al	la Va	al Pl		al Gly L5	
	ata Ile															1897
	ata Ile												-		ttg Leu	1945
caa Gln	gtg Val 50	tgg Trp	atc Ile	ccc Pro	ccc Pro	ctc Leu 55	aac Asn	gtt Val	cgg Arg	GJÅ 333	Gly 60	cgc Arg	gat Asp	gcc Ala	gtc Val	1993
	ctc Leu								_							2041
	atc Ile															2089
	acc Thr													_	gca Ala	2137
	atg Met														_	2185
	Met 130															2233
	cca Pro															2281
	gtt Val															2329
	GJA 333													_		2377
	tcc Ser											_		_		2425
	gaa Glu 210	_										•	_		•	2473
	cag Gln	_										-				2521

													tcc Ser		gca Ala	2569
							•	=					tgg Trp 270		gtc Val	2617
	_		_			_			_	_		_	ggc Gly			2665
							_	_				_	tgg Trp		gcg Ala	2713
		_		_		_				_			agc Ser		_	2761
				_	• •	_	_	_		_			cgc Arg		cgg Arg	2809
													tcc Ser 350		ttg Leu	2857
_			=		·		_		_				cac		gtg Val	2905
	_	_	•		· _	•	_	_	-	· _		_			gtg Val	2953
_							_	_			_		tcc Ser	_	_	3001
					Ser	Pro	Pro	Ala	Val	Pro		Thr	Phe	_	gtg Val	3049
													aag Lys 430		ccg Pro	3097
								_					aac Asn			3145
												-	gca Ala			3193

	_				aga Arg 470			_				_	 	3241
	_				acc Thr		_	_	_		_		 =	3289
					gac Asp							_		3337
	Ser				ctg Leu								gag Glu	3385
	_	_			ctc Leu	_	_						gga Gly	3433
					cat His 550									3481
					ttt Phe								atc Ile	3529
					ctc Leu								gat Asp	3577
_					ctg Leu		_					_	 tat Tyr	3625
				-	gta Val		_			_		_	 att Ile	3673
					gct Ala 630								gac Asp 640	3721
			_	_	aat Asn		_	_			_		_	3769
					acc Thr								gcg Ala	3817
_		_	Ser		cgg Arg		_				·			3865

		agg Arg													gat Asp	3913
		gtt Val													gag Glu 720	3961
ctc Leu	acg Thr	ccc Pro	gcc Ala	gag Glu 725	acc Thr	tca Ser	gtt Val	agg Arg	ttg Leu 730	cgg Arg	gct Ala	tac Tyr	cta Leu	aac Asn 735	aca Thr	4009
		ttg Leu														4057
		ggc Gly 755													aag Lys	4105
		gga Gly													gtg Val	4153
tgc Cys 785	gcc	agg Arg	gct Ala	cag Gln	gct Ala 790	cca Pro	cct Pro	cca Pro	tcg Ser	tgg Trp 795	gac Asp	caa Gln	atg Met	tgg Trp	aag Lys 800	4201
		ata Ile													ctg Leu	4249
		ctg Leu													ata Ile	4297
acc Thr	aaa Lys	tac Tyr 835	atc Ile	atg Met	gca Ala	tgc Cys	atg Met 840	tcg Ser	gct Ala	ġac Asp	ctg Leu	gag Glu 845	gtc Val	gtc Val	acg Thr	4345
		tgg Trp														4393
tgc Cys 865	ctg Leu	aca Thr	aca Thr	ggc Gly	agc Ser 870	gtg Val	gtc Val	att Ile	gtg Val	ggc Gly 875	agg Arg	atc Ile	atc Ile	ttg Leu	tcc Ser 880	4441
gga Gly	aag Lys	ccg Pro	gcc Ala	atc Ile 885	att Ile	ccc Pro	gac Asp	agg Arg	gaa Glu 890	gtc Val	ctt Leu	tac Tyr	Arg	gag Glu 895	ttc Phe	4489
gat Asp		atg Met														4537

	cag Gln														caa Gln	4585
	gcc Ala 930														aag Lys	4633
tgg Trp 945	cgg	acc Thr	ctc Leu	gaa Glu	gcc Ala 950	ttc Phe	tgg Trp	gcg Ala	aag Lys	cat His 955	atg Met	tgg Trp	aat Asn	ttc Phe	atc Ile 960	4681
	gly														ccc Pro	4729
Ala	ata Ile	Ala	Ser 980	Leu	Met	Ala	Phe	Thr 985	Ala	Ser	Ile	Thr	Ser 990	Pro	Leu	4777
	acc		His					Asn					\mathtt{Trp}			4825
	caa Gln 1010	Leu					Ala					Val				4873
Ile 1025	Ala	Gly	Ala	Ala	Val 1030	Gly	Ser	Ile	Gly	Leu 1035	Gly	Lys	Val	Leu	1040	4921
Asp	Ile	Leu	Ala	Gly 1045	Tyr	Gly	Ala	Gly	Val 1050	Ala)	Gly	Ala	Leu	Val 1055		4969
Phe	aag Lys	Val	Met 1060	Ser	Gly	Glu	Met	Pro 1065	Ser	Thr	Glu	Asp	Leu 1070	Val	Asn	5017
cta Leu	ctc Leu	cct Pro 1075	Ala	Ile	ctc Leu	Ser	Pro	ggc Gly	Ala	Leu	Val	Val	Gly	gtc Val	gtg Val	5065
tgc Cys	gca Ala 1090	Ala	ata Ile	ctg Leu	cgt Arg	cgg Arg 1095	His	gtg Val	ggc Gly	cca Pro	999 Gly 1100	Glu	gly aaa	gct Ala	gtg Val	5113
cag Gln 1105	Trp	atg Met	aac Asn	cgg Arg	ctg Leu 1110	Ile	gcg Ala	ttc Phe	gct Ala	tcg Ser 1115	Arg	ggt Gly	aac Asn	cac His	gtc Val 1120	5161
tcc Ser	CCC	acg	cac	tat	gtg	cct	gag	agc	gac	gct	gca	gca	cgt	gtc	act	ن ⁵ 209

78/93

cag atc ctc tct ag Gln Ile Leu Ser Se 1140		Gln Leu Leu Lys	
cag tgg atc aac ga Gln Trp Ile Asn Gl 1155	- -		Ser Trp Leu
aga gat gtt tgg ga Arg Asp Val Trp As 1170			
tgg ctc cag tcc aa Trp Leu Gln Ser Ly 1185	_		ccc ttc ttc 5401 Pro Phe Phe 1200
tca tgt caa cgt gg Ser Cys Gln Arg Gl 12	_		
caa acc acc tgc co Gln Thr Thr Cys Pr 1220		Ile Thr Gly His	
ggt tcc atg agg at Gly Ser Met Arg Il 1235	_	_	Thr Trp His
gga aca ttc ccc at Gly Thr Phe Pro Il 1250			_
ccg gcg cca aat ta Pro Ala Pro Asn Ty 1265			-
tac gtg gag gtt ac Tyr Val Glu Val Th	-		— — — <u>—</u>
acc act gac aac gt Thr Thr Asp Asn Va 1300		Gln Val Pro Ala	_
ttc aca gaa gtg ga Phe Thr Glu Val As 1315			Pro Ala Cys
aaa ccc ctc cta cg Lys Pro Leu Leu Ar 1330		-	
tac ctg gtt ggg to Tyr Leu Val Gly Se			

gtg Val	ctc Leu	act Thr	tcc Ser	atg Met 136	Leu	acc Thr	gac Asp	ccc Pro	tcc Ser 137	His	att Ile	acg Thr	gcg Ala	gag Glu 137		5929
gct Ala	aag Lys	cgt Arg	agg Arg 138	Leu	gcc Ala	agg Arg	gga Gly	tct Ser 138	Pro	ccc	tcc Ser	ttg Leu	gcc Ala 139	Ser	tca Ser	5977
tca Ser	gct Ala	agc Ser 139	Gln	ctg Leu	tct Ser	gcg Ala	cct Pro 140	Ser	ttg Leu	aag Lys	gca Ala	aca Thr 140	Cys	act Thr	acc Thr	6025
cgt Arg	cat His 141	Asp	tcc Ser	ccg Pro	gac Asp	gct Ala 141	Asp	ctc Leu	atc Ile	gag Glu	gcc Ala 1420	Asn	ctc Leu	ctg Leu	tgg Trp	6073
cgg Arg 142	Gln	gag Glu	atg Met	gly	999 Gly 1430	Asn	atc Ile	acc Thr	cgc Arg	gtg Val 143	Glu	tca Ser	gaa Glu	aat Asn	aag Lys 1440	6121
gta Val	gta Val	att Ile	Leu	gac Asp 1445	Ser	ttc Phe	gag Glu	ccg Pro	ctc Leu 1450	Gln	gcg Ala	gag Glu	gag Glu	gat Asp 1455	Glu	6169
agg Arg	gaa Glu	gta Val	tcc Ser 1460	Val	ccg Pro	gcg Ala	gag Glu	atc Ile 1465	Leu	Arg	agg Arg	tcc Ser	agg Arg 1470	Lys	ttc Phe	6217
cct Pro	cga Arg	gcg Ala 1475	Met	ccc Pro	ata Ile	tgg. Trp	gca Ala 1480	Arg	ccg Pro	gat Asp	tac Tyr	aac Asn 1485	Pro	cca Pro	ctg Leu	6265
tta Leu	gag Glu 1490	tcc Ser	tgg Trp	aag Lys	gac Asp	ccg Pro 1495	Asp	tac Tyr	gtc Val	cct Pro	cca Pro 1500	Val	gta Val	cac His	gjå aaa	6313
tgt Cys 150!	Pro	ttg Leu	ccg Pro	cct Pro	gcc Ala 1510	Lys	gcc Ala	cct Pro	ccg Pro	ata Ile 1515	Pro	cct Pro	cca Pro	cgg Arg	agg Arg 1520	6361
aag Lys	agg Arg	acg Thr	gtt Val	gtc Val 1525	Leu	Ser	Glu	tct Ser	Thr	Val	tct Ser	tct Ser	gcc Ala	ttg Leu 1535	Ala	6409
gag Glu	ctc Leu	gcc Ala	aca Thr 1540	Lys	acc Thr	ttc Phe	Gly	agc Ser 1545	Ser	gaa Glu	tcg Ser	tcg Ser	gcc Ala 1550	Val	gac Asp	6457
agc Ser	ggc	acg Thr 1555	Ala	acg Thr	gcc Ala	Ser	cct Pro 1560	Asp	cag Gln	ccc Pro	tcc Ser	gac Asp 1565	Asp	ggc Gly	gac Asp	6505
gcg Ala	gga Gly 1570	tcc Ser	gac Asp	gtt Val	gag Glu	tcg Ser 1575	Tyx	tcc Ser	tcc Ser	atg Met	ccc Pro 1580	Pro	ctt Leu	gag Glu	ej aaa	6553

		<u></u>	tgg tct acc gta agc 6601 Trp Ser Thr Val Ser 1600
			atg tcc tac aca tgg 6649 Met Ser Tyr Thr Trp 1615
	Ile Thr Pro		gaa acc aag ctg ccc 6697 Glu Thr Lys Leu Pro 1630
	Ser Asn Ser	_	cac aac ttg gtc tat 6745 His Asn Leu Val Tyr 1645
	_	Ser Leu Arg Gln	aag aag gtc acc ttt 6793 Lys Lys Val Thr Phe 1660
_	_	· ·	gac gtg ctc aag gag 6841 Asp Val Leu Lys Glu 1680
		-	ctt cta tcc gtg gag 6889 Leu Leu Ser Val. Glu 1695
	Leu Thr Pro		aga tct aaa ttt ggc 6937 Arg Ser Lys Phe Gly 1710
		_	aag gcc gtt aac cac 6985 Lys Ala Val Asn His 1725
		Leu Leu Glu Asp	act gag aca cca att 7033 Thr Glu Thr Pro Ile 1740
_	Met Ala Lys	Asn Glu Val Phe	tgc gtc caa cca gag 7081 Cys Val Gln Pro Glu 5 1760
			ttc cca gat ttg ggg 7129 Phe Pro Asp Leu Gly 1775
<u> </u>	Glu Lys Met	-	gtg gtc tcc acc ctc 7177 Val Val Ser Thr Leu 1790
cct cag gcc gtg Pro Gln Ala Val			caa tac tct cct gga 7225

		Val					Asn					Lys		_	cct Pro	7273	
	Gly	_				Thr			_	_	Ser	acg Thr	_		<i>-</i>	7321	•
					Glu					Gln		tgt Cys		_	Ala	7369	
				Gln					Leu			cgg Arg		Tyr		7417	
Gly	Gly	Pro 1875	Leu	Thr	Asn	Ser	Lys 1880	Gly	Gln	Asn	Cys	ggc Gly 1885	Tyr	Arg	Arg	7465	
Cys	Arg 1890	Ala	Ser	Gly	Val	Leu 1895	Thr	Thr	Ser	Cys	Gly 1900	Asn)	Thr	Leu		7513	
Cys 1905	Tyr	Leu	Lys	Ala	Ala 1910	Ala)	Ala	Cys	Arg	Ala 1915	Ala	aag Lys	Leu	Gln	Asp 1920	7561	
					Cys					Val		atc Ile	-	-	Ser	7609 _.	
Ala	Gly	Thr	Gln 1940	Glu	Asp	Glu	Ala	Ser 1945	Leu	Arg	Ala	ttc Phe	Thr 1950	Glu	Ala	7657	
Met	Thr	Arg 1955	Tyr	Ser	Ala	Pro	Pro 1960	Gly	Asp	Pro	Pro	aaa Lys 1965	Pro	Glu	Tyr	7705	
Asp	Leu 1970	Glu	Leu	Ile	Thr	Ser 1975	Cys	Ser	Ser	Asn	Val 1980		Val	Ala	His	7753	
	Ala					Val			Leu		Arg	gac Asp		Thr		7801	
					Ala		_	Thr	_	Arg		act Thr	Pro	_	Asn	7849	
				Asn			Met		Ala			ttg Leu		Ala		7897	

82/93

atg atc ctg atg act cat ttc ttc tcc atc ctt cta gct cag gaa caa 7 Met Ile Leu Met Thr His Phe Phe Ser Ile Leu Leu Ala Gln Glu Gln 2035 2040 2045	7945
ctt gaa aaa gcc cta gat tgt cag atc tac ggg gcc tgt tac tcc att Leu Glu Lys Ala Leu Asp Cys Gln Ile Tyr Gly Ala Cys Tyr Ser Ile 2050 2055 2060	7993
gag cca ctt gac cta cct cag atc att caa cga ctc cac ggc ctt agc 8 Glu Pro Leu Asp Leu Pro Gln Ile Ile Gln Arg Leu His Gly Leu Ser 2065 2070 2075 2080	3041
gca ttt tca ctc cat agt tac tct cca ggt gag atc aat agg gtg gct Ala Phe Ser Leu His Ser Tyr Ser Pro Gly Glu Ile Asn Arg Val Ala 2085 2090 2095	8089
tca tgc ctc agg aaa ctt ggg gta ccg ccc ttg cga gtc tgg aga cat Ser Cys Leu Arg Lys Leu Gly Val Pro Pro Leu Arg Val Trp Arg His 2100 2105 2110	8137
cgg gcc aga agt gtc cgc gct agg cta ctg tcc cag ggg ggg agg gct Arg Ala Arg Ser Val Arg Ala Arg Leu Leu Ser Gln Gly Gly Arg Ala 2125	8185
gcc act tgt ggc aag tac ctc ttc aac tgg gca gta agg acc aag ctc Ala Thr Cys Gly Lys Tyr Leu Phe Asn Trp Ala Val Arg Thr Lys Leu 2130 2135 2140	8233
aaa ctc act cca atc ccg gct gcg tcc cag ttg gat tta tcc agc tgg Lys Leu Thr Pro Ile Pro Ala Ala Ser Gln Leu Asp Leu Ser Ser Trp 2145 2150 2155 2160	8281
ttc gtt gct ggt tac agc ggg gga gac ata tat cac agc ctg tct cgt Phe Val Ala Gly Tyr Ser Gly Gly Asp Ile Tyr His Ser Leu Ser Arg 2165 2170 2175	8329 _.
gcc cga ccc cgc tgg ttc atg tgg tgc cta ctc cta ctt tct gta ggg Ala Arg Pro Arg Trp Phe Met Trp Cys Leu Leu Leu Ser Val Gly 2180 2185 2190	8377
gta ggc atc tat cta ctc ccc aac cga tga acggggagct aaacactcca Val Gly Ile Tyr Leu Leu Pro Asn Arg * 2195 2200	8427
ggccaatagg ccatcctgtt tttttccctt ttttttttc tttttttt ttttttt	8547
<210> 25 <211> 8638 <212> DNA <213> HCV	
<220> <221> CDS	

<222> (1802)...(8407)

```
<400> 25
accageceee gattggggge gacaeteeae catagateae teeeetgtga ggaaetaetg 60
tcttcacgca gaaagcgtct agccatggcg ttagtatgag tgtcgtgcag cctccaggac 120
ccccctccc gggagagcca tagtggtctg cggaaccggt gagtacaccg gaattgccag 180
gacgaccggg teetttettg gateaacccg etcaatgeet ggagatttgg gegtgeece 240
gcgagactgc tagccgagta gtgttgggtc gcgaaaggcc ttgtggtact gcctgatagg 300
gtgcttgcga gtgccccggg aggtctcgta gaccgtgcac catgagcacg aatcctaaac 360
ctcaaagaaa aaccaaaggg cgcgccatga ttgaacaaga tggattgcac gcaggttctc 420
cggccgcttg ggtggagagg ctattcggct atgactgggc acaacagaca atcggctgct 480
ctgatgccgc cgtgttccgg ctgtcagcgc aggggcgccc ggttcttttt gtcaagaccg 540
acctgtccgg tgccctgaat gaactgcagg acgaggcagc gcggctatcg tggctggcca 600
cgacgggcgt tccttgcgca gctgtgctcg acgttgtcac tgaagcggga agggactggc 660
tgctattggg cgaagtgccg gggcaggatc tcctgtcatc tcaccttgct cctgccgaga 720
aagtatecat catggetgat geaatgegge ggetgeatae gettgateeg getaeetgee 780
cattegacea ceaagegaaa eategeateg agegageaeg taeteggatg gaageeggte 840
ttgtcgatca ggatgatctg gacgaagagc atcaggggct cgcgccagcc gaactgttcg 900
ccaggctcaa ggcgcgcatg cccgacggcg aggatctcgt cgtgacccat ggcgatgcct 960
gcttgccgaa tatcatggtg gaaaatggcc gcttttctgg attcatcgac tgtggccggc 1020
tgggtgtggc ggaccgctat caggacatag cgttggctac ccgtgatatt gctgaagagc 1080
ttggcggcga atgggctgac cgcttcctcg tgctttacgg tatcgccgct cccgattcgc 1140
agegeatege ettetatege ettettgaeg agttettetg agttegegee eagatgttaa 1200
cagaccacaa cggtttccct ctagcgggat caattccgcc ccccccta acgttactgg 1260
ccgaagccgc ttggaataag gccggtgtgc gtttgtctat atgttatttt ccaccatatt 1320
geegtetttt ggeaatgtga gggeeeggaa acetggeeet gtettettga egageattee 1380
taggggtctt tcccctctcg ccaaaggaat gcaaggtctg ttgaatgtcg tgaaggaagc 1440
agttcctctg gaagcttctt gaagacaaac aacgtctgta gcgacccttt gcaggcagcg 1500
gaacccccca cctggcgaca ggtgcctctg cggccaaaag ccacgtgtat aagatacacc 1560 ·
tgcaaaggcg gcacaacccc agtgccacgt tgtgagttgg atagttgtgg aaagagtcaa 1620
atggctctcc tcaagcgtat tcaacaaggg gctgaaggat gcccagaagg taccccattg 1680
tatgggatct gatctggggc ctcggtgcac atgctttaca tgtgtttagt cgaggttaaa 1740
aaacgtctag gccccccgaa ccacggggac gtggttttcc tttgaaaaac acgataatac 1800
c atg gac cgg gag atg gca gca tcg tgc gga ggc gcg gtt ttc gta ggt 1849
  Met Asp Arg Glu Met Ala Ala Ser Cys Gly Gly Ala Val Phe Val Gly
   1
                                       10
                   5
                                                           15
ctg ata ctc ttg acc ttg tca ccg cac tat aag ctg ttc ctc gct agg
                                                                   1897
Leu Ile Leu Leu Thr Leu Ser Pro His Tyr Lys Leu Phe Leu Ala Arg
             20
                                 25
                                                      30
ctc ata tgg tgg tta caa tat ttt atc acc agg gcc gag gca cac ttg
                                                                   1945
Leu Ile Trp Trp Leu Gln Tyr Phe Ile Thr Arg Ala Glu Ala His Leu
         35
                             40
                                                  45
caa gtg tgg atc ccc ccc ctc aac gtt cgg ggg ggc cgc gat gcc gtc
                                                                   1993
Gln Val Trp Ile Pro Pro Leu Asn Val Arg Gly Gly Arg Asp Ala Val
     50
                         55
                                             60
atc ctc ctc acg tgc gcg atc cac cca gag cta atc ttt acc atc acc
                                                                   2041
Ile Leu Leu Thr Cys Ala Ile His Pro Glu Leu Ile Phe Thr Ile Thr
 65
                     70
                                         75
                                                             80
aaa atc ttg ctc gcc ata ctc ggt cca ctc atg gtg ctc cag gct ggt
                                                                   2089
Lys Ile Leu Leu Ala Ile Leu Gly Pro Leu Met Val Leu Gln Ala Gly
                 85
                                                         95
                                     90
ata acc aaa gtg ccg tac ttc gtg cgc gca cac ggg ctc att cgt gca
                                                                   2137
Ile Thr Lys Val Pro Tyr Phe Val Arg Ala His Gly Leu Ile Arg Ala
            100
                                105
                                                     110
```

					aag Lys										gct Ala	2185
					gca Ala										ctc Leu	2233
					tgg Trp 150										gtg Val 160	2281
					gtc Val										acc Thr	2329
					gcg Ala										ccc Pro	2377
gtc Val	tcc Ser	gcc Ala 195	cgc Arg	agg Arg	GJÀ aaa	agg Arg	gag Glu 200	ata Ile	cat His	ctg Leu	gga Gly	ccg Pro 205	gca Ala	gac Asp	agc Ser	2425
					tgg Trp	-			_						tcc Ser	2473
					cta Leu 230										ggc Gly 240	2521
Arg	Asp	Arg	Asn	Gln 245	Val	Glu	Gly	Glu	Val 250	Gln	Val	Val	Ser	Thr 255		2569
					gcg										gtc Val	2617
			Ala	Gly	tca Ser	Lys	Thr	Leu	Ala		Pro					2665
					aat Asn										gcg Ala	2713
					tcc Ser 310											2761
					aag Lys											2809

	_			agg Arg 340		_										ttg Leu	2857
				tcg Ser												gtg Val	2905
				Arg												gtg Val	2953
				ccc Pro												gtc Val 400	3001
				aac Asn												gtg Val	3049
				cac His 420												ccg Pro	3097
				gca Ala												tcc Ser	3145
	gtc Val	gcc Ala 450	gcc Ala	acc Thr	cta Leu	ggt Gly	ttc Phe 455	Gly 999	gcg Ala	tat Tyr	atg Met	tct Ser 460	aag Lys	gca Ala	cat His	ggt Gly	3193
				aac Asn												gcc Ala 480	3241
				tac Tyr												tgc Cys	3289
	tct Ser	Gly 999	ggc	gcc Ala 500	tat Tyr	gac Asp	atc Ile	ata Ile	ata Ile 505	tgt Cys	gat Asp	gag Glu	tgc Cys	cac His 510	tca Ser	act Thr	3337
				act Thr													3385
•				gcg Ala											_		3433
				gtg Val										_		_	3481

					ttt Phe										atc Ile	3529
					ctc Leu											3577
					ctg Leu										tat Tyr	3625
					gta Val										att Ile	3673
					gct Ala 630										gac Asp 640	3721
					aat Asn										agc Ser	3769
ctg Leu	gac Asp	ccg Pro	acc Thr 660	ttc Phe	acc Thr	att Ile	gag Glu	acg Thr 665	acg Thr	acc Thr	gtg Val	cca Pro	caa Gln 670	gac Asp	gcg Ala	3817
					cgg Arg										ggc Gly	3865
att Ile	tac Tyr 690	agg Arg	ttt Phe	gtg Val	act Thr	cca Pro 695	gga Gly	gaa Glu	cgg Arg	ccc Pro	tcg Ser 700	Gly	atg Met	ttc Phe	gat Asp	3913
					gag Glu 710										gag Glu 720	3961
ctc Leu	acg Thr	ccc Pro	gcc Ala	gag Glu 725	acc Thr	tca Ser	gtt Val	agg Arg	ttg Leu 730	cgg Arg	gct Ala	tac Tyr	cta Leu	aac Asn 735	aca Thr	4009
												Trp			gtc Val	4057
					cac His											4105
					ttc Phe											4153
					gct Ala 790											4201

					aag Lys										ctg Leu	4249
				_	gtt Val											4297
					gca Ala										acg Thr	4345
					gta Val											4393
tgc Cys 865	ctg Leu	aca Thr	aca Thr	Gly	agc Ser 870	gtg Val	gtc Val	att Ile	gtg Val	ggc Gly 875	agg Arg	atc Ile	atc Ile	ttg Leu	tcc Ser 880	4441
gga Gly	agg Arg	ccg Pro	gcc Ala	atc Ile 885	att Ile	ccc Pro	gac Asp	agg Arg	gaa Glu 890	gtc Val	ctt Leu	tac Tyr	cgg Arg	gag Glu 895	ttc Phe	4489
					tgt Cys											4537
atg Met	cag Gln	ctc Leu 915	gcc Ala	gaa Glu	caa Gln	ttc Phe	aaa Lys 920	cag Gln	aag Lys	gca Ala	atc Ile	ggg Gly 925	ttg Leu	ctg Leu	caa Gln	4585
Thr	Ala 930	Thr	Lys	Gln	gcg Ala	Glu 935	Ala	Ala	Ala	Pro	Val 940	Val	Glu	Ser	Lys	4633
Trp 945	Arg	Thr	Leu	Glu	gcc Ala 950	Phe	Trp	Ala.	Lys	His 955	Met	Trp	Asn	Phe	Ile 960	4681
agc Ser	GJA aaa	ata Ile	caa Gln	tat Tyr 965	tta Leu	gca Ala	Gly	Leu	tcc Ser 970	Thr	ctg Leu	cct Pro	Gly	aac Asn 975	ccc Pro	4729
					atg Met										ctc Leu	47,77
acc Thr	acc Thr	caa Gln 995	His	acc Thr	ctc Leu	ctg Leu	ttt Phe 1000	Asn	atc Ile	ctg Leu	gjå aaa	gga Gly 1005	Trp	gtg Val	gcc Ala	4825
gcc Ala	caa Gln 1010	Leu	gct Ala	cct Pro	ccc Pro	agc Ser 1015	Ala	gct Ala	tcc Ser	gct Ala	ttc Phe 1020	Val	Gly	gcc Ala	ggc [.]	4873

atc gct gga gcg Ile Ala Gly Ala 1025				-	4921
gat att ttg gca Asp Ile Leu Ala			Ala Gly Ala		4969
ttt aag gtc atg Phe Lys Val Met 1060	Ser Gly Glu				5017
cta ctc cct gct Leu Leu Pro Ala 1075				Gly Val Val	5065
tgc gca gcg ata Cys Ala Ala Ile 1090		His Val Gly			5113
cag tgg atg aac Gln Trp Met Asn 1105					5161
tcc ccc acg cac Ser Pro Thr His			Ala Ala Ala		5209
cag atc ctc tct Gln Ile Leu Ser 1140	Ser Leu Thr				5257
cag tgg atc aac Gln Trp Ile Asn 1155				Ser Trp Leu	5305
aga gat gtt tgg Arg Asp Val Trp 1170		Cys Thr Val		_	5353
tgg ctc cag tcc Trp Leu Gln Ser 1185	Lys Leu Leu				5401
tca tgt caa cgt Ser Cys Gln Arg			Arg Gly Asp		5449
caa acc acc tgc Gln Thr Thr Cys 1220	Pro Cys Gly				5497
tgt tcc atg agg Cys Ser Met Arg 1235					5545

Gly Thr Phe 1250		- -	acc acg ggc Thr Thr Gly			5593
_	Asn Tyr Se		ctg tgg cgg Leu Trp Arg 1275	Val Ala Ala		5641
_			gat ttc cac Asp Phe His 1290			5689
		s Cys Pro (tgt cag gtt Cys Gln Val 1305	-	Glu Phe	5737
	Val Asp Gl	· ·	ttg cac agg Leu His Arg	-	- -	5785
			aca ttc ctg Thr Phe Leu	- - - -		5833
	Gly Ser Gl		tgc gag ccc Cys Glu Pro 1355	Glu Leu Asp		5881
			ccc tcc cac Pro Ser His 1370		- -	5929
gct aag-cgt	agg ctg gc	c agg gga t	tet eee eee	tee tta gee	agc tca	5977
Ala Lys Arg	Arg Leu Al 1380		Ser Pro Pro 1385		Ser Ser	
tca gct agc	cag ctg to	t gcg cct t	Ser Pro Pro 1385 tcc ttg aag Ser Leu Lys	Ser Leu Ala 139 gca aca tgc	Ser Ser 0 act acc	6025
tca gct agc Ser Ala Ser 139 cgt cat gac Arg His Asp	cag ctg to Gln Leu Se tcc ccg ga	t gcg cct i r Ala Pro i 1400 c gct gac i p Ala Asp i	Ser Pro Pro 1385 tcc ttg aag Ser Leu Lys ctc atc gag Leu Ile Glu	Ser Leu Ala 139 gca aca tgc Ala Thr Cys 1405 gcc aac ctc	Ser Ser 0 act acc Thr Thr ctg tgg	
tca gct agc Ser Ala Ser 139 cgt cat gac Arg His Asp 1410 cgg cag gag	cag ctg to Gln Leu Se Ser Ccg ga Ser Pro As	t gcg cct is r Ala Pro is 1400 c gct gac of p Ala Asp is 1415	Ser Pro Pro 1385 tcc ttg aag Ser Leu Lys ctc atc gag Leu Ile Glu	Ser Leu Ala 139 gca aca tgc Ala Thr Cys 1405 gcc aac ctc Ala Asn Leu 1420 gag tca gag Glu Ser Glu	Ser Ser o act acc Thr Thr ctg tgg Leu Trp aat aag	6025
tca gct agc Ser Ala Ser 139 cgt cat gac Arg His Asp 1410 cgg cag gag Arg Gln Glu 1425 gta gta att	cag ctg to Gln Leu Se tcc ccg ga Ser Pro As atg ggc gg Met Gly Gl 14	t gcg cct is 1400 c gct gac op Ala Asp 1415 g aac atc ay Asn Ile is 30	Ser Pro Pro 1385 tcc ttg aag Ser Leu Lys ctc atc gag Leu Ile Glu acc cgc gtg Thr Arg Val	Ser Leu Ala 139 gca aca tgc Ala Thr Cys 1405 gcc aac ctc Ala Asn Leu 1420 gag tca gag Glu Ser Glu gcg gag gag	Ser Ser o act acc Thr Thr ctg tgg Leu Trp aat aag Asn Lys 1440 gat gag	6025

-		_	gat tac aac cct Asp Tyr Asn Pro 1485	
tta gag tcc tgg Leu Glu Ser Trp 1490		Asp Tyr Val	cct cca gtg gta Pro Pro Val Val 1500	
			ata cca cct cca Ile Pro Pro Pro 1515	
aag agg acg gtt Lys Arg Thr Val				
	Lys Thr Phe		gaa tcg tcg gcc Glu Ser Ser Ala 155	Val Asp
			ccc tcc gac gac Pro Ser Asp Asp 1565	
		Tyr Ser Ser	atg ccc ccc ctt Met Pro Pro Leu 1580	
			tct tgg tct acc Ser Trp Ser Thr 1595	
			tcg atg tcc tac Ser Met Ser Tyr O	_
	Ile Thr Pro		gag gaa acc aag Glu Glu Thr Lys	Leu Pro
			163	U
Ile Asn Ala Leu	_	Leu Leu Arg	cac cac aac ttg His His Asn Leu	gtc tat 6745
Ile Asn Ala Leu 1635 gct aca aca tct	Ser Asn Ser	Leu Leu Arg 1640 agc ctg cgg Ser Leu Arg	cac cac aac ttg His His Asn Leu	gtc tat 6745 Val Tyr acc ttt 6793
Ile Asn Ala Leu 1635 gct aca aca tct Ala Thr Thr Ser 1650 gac aga ctg cag	ser Asn Ser cgc agc gca Arg Ser Ala 1659 gtc ctg gac	Leu Leu Arg 1640 agc ctg cgg Ser Leu Arg 5	cac cac aac ttg His His Asn Leu 1645 cag aag aag gtc Gln Lys Lys Val	gtc tat 6745 Val Tyr acc ttt 6793 Thr Phe aag gag 6841

		_			Leu		ccc Pro		_	Ser					Phe	Gly	6937
				Lys			arg		Leu					Val		cac His	6985
			Ser				gac Asp 1735	Leu					Glu				7033
		Thr					Lys					Cys				gag Glu 1760	7081
						Pro	gct Ala				Val					Gly	7129
					Glu		atg Met			Tyr					Thr	ctc Leu	7177
				Val			tct Ser		Tyr					Ser			7225
	Gln		Val				gtg Val 1815	Asn					Lys				7273
		Gly					acc Thr					Ser					7321
						Glu	gag Glu				Gln					Ala	7369
					Gln	Ala	ata Ile	Arg	Ser	Leu	Thr		Arg		Tyr		7417
		Gly		Leu			Ser		Gly					Tyr		cgg Arg	7465
	Сув		Ala				ctg Leu 1895	Thr					Asn				7513
• (Tyr					gcg Ala					Ala					7561

					Cys					Val.				gaa Glu 1935	Ser	7609
				Glu					Leu				_	gag Glu O	_	7657
			Tyr					Gly					Pro	gaa Glu		7705
		Glu					Cys					Ser		gcg Ala		7 753
	Ala		_			Val				_	Arg	-		acc Thr		7801
					Ala					Arg				gtc Val 2015	Asn	7849
			Gly						Ala					gca Ala)		7897
-		-	Met					Ser					Gln	gaa Glu		7945
		Lys					Gln					Cys		tcc Ser		7993
	Pro					Gln					Leu			ctt Leu	agc Ser 2080	8041
				His	Ser	Tyr	Ser	Pro	Gly	Glu	Ile	Asn	Arg	gtg Val 2095	Ala	8089
				Lys		_			Pro	_		_		aga Arg)		8137
~~~																
_	_		Ser	_		gct Ala		Leu			_		Gly	agg Arg	_ <del>-</del>	8185

WO 02/052015

4 9

#### 93/93

	Leu				_	gct Ala )			_	_	Asp			_		8281
					Ser	gjy aaa				Tyr	_				Arg	8329
_	_		_	Trp		atg Met		_	Leu					Val		8377
		_	Tyr			Pro		Arg	_	acgg	ggag	gct a	aaaca	actċo		8427
tttt	tttt ctta	ett t age o	tttt cctac	etect gtcad	t tt	tttt	tcct	ctt gaaa	tttt	tcc	ttt	tcttt	cc t	ttgg	ttttt gtggct cagaga	8547

·•

# (19) World Intellectual Property Organization International Bureau



# 

(43) International Publication Date 4 July 2002 (04.07.2002)

**PCT** 

# (10) International Publication Number WO 02/052015 A3

- (51) International Patent Classification⁷: C12N 15/51, 15/40, C12Q 1/68, 1/70, C12N 5/10, 7/04, 15/85
- (21) International Application Number: PCT/CA01/01843
- (22) International Filing Date:

20 December 2001 (20.12.2001)

(25) Filing Language:

English

(26) Publication Language:

English

US

- (30) Priority Data: 60/257,857 22 December 2000 (22.12.2000)
- (71) Applicant (for all designated States except US): BOEHRINGER INGELHEIM (CANADA) LTD. [CA/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): KUKOLJ, George [CA/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA). PAUSE, Arnim [DE/CA]; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).
- (74) Agent: BERNIER, Louise, G.; 2100 Cunard Street, Laval, Québec H7S 2G5 (CA).

- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

#### Published:

- with international search report
- (88) Date of publication of the international search report: 20 November 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

A3

(54) Title: SELF-REPLICATING RNA MOLECULE FROM HEPATITIS C VIRUS

(57) Abstract: A unique HCV RNA molecule is provided having an enhanced efficiency of establishing cell culture replication. Novel adaptive mutations have been identified within the HCV non-structural region that improves the efficiency of establishing persistently replicating HCV RNA in cell culture. This self-replicating polynucleotide molecule contains, contrary to all previous reports, a 5'-NTR that can be either an A as an alternative to the G already disclosed and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule. The G-->A mutation gives rise to HCV RNA molecules that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, possess greater efficiency of transduction and/or replication. These RNA molecules when transfected in a cell line are useful for evaluating potential inhibitors of HCV replication.

International Application No PCT/CA 01/01843

A. CLASSIFICATION OF SUBJECT MATTER C12Q1/68 C12Q1/70 C12N15/40 C12N15/51 C12N5/10 C12N7/04 C12N15/85 According to International Patent Classification (IPC) or to both national classification and IPC **B. FIELDS SEARCHED** Minimum documentation searched (classification system followed by classification symbols) C12N C12Q IPC 7 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the International search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data, BIOSIS, MEDLINE, SEQUENCE SEARCH C. DOCUMENTS CONSIDERED TO BE RELEVANT Citation of document, with indication, where appropriate, of the relevant passages Category ° Relevant to claim No. LOHMANN V ET AL: "Replication of X 1-22 subgenomic hepatitis C virus RNAs in a hepatoma cell line." SCIENCE (WASHINGTON D C), vol. 285, no. 542, 2 July 1999 (1999-07-02), pages 110-113, XP002232924 ISSN: 0036-8075 the whole document EP 1 043 399 A (BARTENSCHLAGER RALF DR) 1-22 11 October 2000 (2000-10-11) page 3 -page 24; tables 1,3 Further documents are listed in the continuation of box C. Patent family members are listed in annex. Special categories of cited documents: "T" later document published after the International filing date or priority date and not in conflict with the application but "A" document defining the general state of the art which is not cited to understand the principle or theory underlying the considered to be of particular relevance invention "E" earlier document but published on or after the international "X" document of particular relevance; the claimed invention filing date cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another "Y" document of particular relevance; the claimed invention citation or other special reason (as specified) cannot be considered to involve an inventive step when the *O* document referring to an oral disclosure, use, exhibition or document is combined with one or more other such docuother means ments, such combination being obvious to a person skilled in the art. "P" document published prior to the international filing date but "&" document member of the same patent family later than the priority date claimed Date of the actual completion of the international search Date of mailing of the international search report 01 07 2003 3 March 2003 Name and mailing address of the ISA **Authorized officer** European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Schulz, R Fax: (+31-70) 340-3016

Internal Application No PCT/CA 01/01843

C/Continue	etion) DOCUMENTS CONSIDERED TO BE RELEVANT	PC I / CA C	77 020 13
Category °	Citation of document, with indication, where appropriate, of the relevant passages		
omega.,	Oncher of obtaining with inducation, where appropriate, or the relevant passages		Relevant to claim No.
A	BLIGHT K J ET AL: "EFFICIENT INITIATION OF HCV RNA REPLICATION IN CELL CULTURE" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, US, vol. 290, 8 December 2000 (2000-12-08), pages 1972-1974, XP002951271 ISSN: 0036-8075 page 1972 -page 1973; table 1		1-22
PCT//SA/210 (r	continuation of second sheet) (July 1992)		



Boxi	Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)
This Inter	national Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
	Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
	Claims Nos.: because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
	Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II C	Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)
This Intern	national Searching Authority found multiple inventions in this international application, as follows:
	see additional sheet
	s all required additional search fees were timely paid by the applicant, this International Search Report covers all earchable claims.
	s all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment f any additional fee.
	is only some of the required additional search fees were timely paid by the applicant, this International Search Report overs only those claims for which fees were paid, specifically claims Nos.:
	to required additional search fees were timely paid by the applicant. Consequently, this International Search Report is estricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark or	The additional search fees were accompanied by the applicant's protest.  No protest accompanied the payment of additional search fees.

#### FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-22

A hepatitis C (HCV) replicon comprising: a 5'-non translated region (NTR) wherein guanin at position 1 is substituted for adenine, a HCV polyprotein region coding for a HCV polyprotein further comprising one or more amino acid substitutions (adaptive mutations) in a non-structural protein and a 3'NTR; a eurkaryotic host cell transfected with said replicon; a RNA replication assay making use of said host cell and a method for testing compounds that inhibit HCV replication using said host cell.

2. Claims: 23-42 (in part)

A hepatitis C (HCV) replicon comprising: a 5'-non translated region (NTR), a HCV polyprotein region coding for a HCV polyprotein comprising a R(1135)K amino acid substitution (adaptive mutation) and a 3'NTR; a eurkaryotic host cell transfected with said replicon; a RNA replication assay making use of said host cell and a method for testing compounds that inhibit HCV replication using said host cell.

3. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a S(1148)G substitution.

4. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a S(1560)G substitution.

5. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a K(1691)R substitution.

6. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a L(1701)F substitution.

7. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a I(1984)V substitution.

#### FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

8. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a T(1993)A substitution.

9. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a G(2042)C or a G(2042)R substitution.

10. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a S(2404)P substitution.

11. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a L(2155)P substitution.

12. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a P(2166)L substitution.

13. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a M(2992)T substitution.

14. Claims: 23-42 (in part)

Idem as subject-matter 2 wherein the adaptive mutation is a E(1202)G substitution.

ormation on patent family members

PCT/CA 01/01843

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1043399	A 11-10-2000	DE 19915178 A1 AT 236988 T AU 2518000 A CA 2303526 A1 DE 50001673 D1 EP 1043399 A2 JP 2001017187 A	05-10-2000 15-04-2003 19-10-2000 03-10-2000 15-05-2003 11-10-2000 23-01-2001

Form PCT/ISA/210 (patent family annex) (July 1992)